

**Signal detection in the auditory midbrain:
Neural correlates and mechanisms
of spatial release from masking**

by

Courtney C. Lane

B. S., Electrical Engineering
Rice University, 1996

SUBMITTED TO THE
HARVARD-MIT DIVISION OF HEALTH SCIENCE AND TECHNOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

OCTOBER 2003

© 2003 Courtney C. Lane. All rights reserved.

The author hereby grants to MIT permission to reproduce
and to distribute publicly paper and electronic
copies of this thesis document in whole or in part.

Signature of Author: _____
Harvard-MIT Division of MIT of Health Sciences and Technology
October 9, 2003

Certified by: _____
Bertrand Delgutte, Ph.D.
Associate Professor of Otology and Laryngology, Harvard Medical School
Senior Research Scientist, RLE, MIT

Accepted by: _____
Martha L. Gray, Ph. D.
Edward Hood Taplin Professor Medical and Electrical Engineering
Co-Director, Harvard-MIT Division of Health Science and Technology

Abstract

Normal-hearing listeners have a remarkable ability to hear in noisy environments, while hearing-impaired listeners and automatic speech-recognition systems often have difficulty in noise. With the ultimate goal of improving hearing aids and speech-recognition systems, we study the neural mechanisms involved in one aspect of noisy-environment listening, “spatial release from masking,” which is the observation that a signal is more easily detected when its source is spatially separated from a masking-noise source. We use neurophysiology, computational modeling, and psychoacoustics to investigate the neural mechanisms of spatial release from masking, and we focus on low frequencies, which are important for speech recognition and are often spared in hearing-impaired listeners. Previous studies suggest that at low frequencies, listeners use interaural time differences (ITDs) to improve signal detection when signals and maskers are spatially separated in azimuth.

To determine how individual neurons respond to spatially separated signals and maskers, we record in anesthetized cats from low-frequency, ITD-sensitive neurons in the inferior colliculus (IC), a major center of converging auditory pathways in the midbrain. We develop a computational model of the neuron responses, which incorporates both interaural cross-correlation (as used in existing binaural models) and amplitude-modulation sensitivity. The need for modulation sensitivity to predict the neural responses indicates that binaural and temporal processing are interacting in signal detection, rather than acting independently as is often assumed. This modification is especially important because most natural sounds, including speech, have pronounced envelope fluctuations that previous models of binaural detection have not utilized.

To relate these neurophysiological results to human behavioral thresholds, we define population thresholds based on the most sensitive neurons in the population. The neural population thresholds are similar to human behavioral thresholds, indicating that low-frequency, ITD-sensitive neurons in the IC may be necessary for low-frequency spatial release from masking in humans. Both interaural correlation and modulation sensitivity seem to be required for the model population thresholds to predict human behavioral thresholds. Overall, our findings suggest that considering the auditory system’s modulation sensitivity and interaural cross-correlation in the design of hearing aids and speech-recognition systems may improve these devices’ performance in noise.

Acknowledgments

As I have worked on this thesis, I have drawn on the help and kindness of many people. Every time I asked for help, someone willingly gave their time, energy, and knowledge. This community, which includes EPL, Boston University, and auditory researchers in general, is a treasure.

I would first like to thank my thesis advisor, Bertrand Delgutte, who has always willingly given his time and attention. His critical and rigorous approach to science has undoubtedly improved every aspect of this project. He never failed to push me to think more deeply, and he allowed me to work independently, giving me full ownership of this thesis. His trust in me has been a cherished gift.

I am also indebted to my thesis committee members. We have had many thought-provoking and lively discussions, and they have proven to be patient and understanding, as well as very fast and thorough readers. Steve Colburn answered my barrage of questions unwearyingly and with a smile. I have taken full advantage of Steve's patience and clarity of thought, and I am especially grateful that he welcomed me into his lab for a summer. John Guinan never failed to offer thoughtful opinions. He served as the outside voice, providing an invaluable perspective and fodder for many interesting discussions. Barb Shinn-Cunningham's ability to see problems and then solve them, regardless of the topic, was invaluable. Her vision, enthusiasm, and humor energized and inspired me.

Also, Norbert Kopco and Barb volunteered their time and grant money for the psychophysical experiments, which has made the thesis far more interesting. Through the long conversations that led to these experiments, I realized just how hard psychophysics is.

EPL is filled with so many knowledgeable people, and I want to thank everyone for providing such a stimulating and welcoming atmosphere. In particular, Nelson Kiang enticed me into the Speech and Hearing program by helping me realize that I wanted to understand how the ear and brain work. We have since had many fascinating conversations on a wide variety of topics, many of which led to advances in this thesis. Joe Adams endured the most ridiculous questions of anyone, and he kindly shared his knowledge of histology and anatomy. Chris Shera taught me that the only way to understand something is to model it; I expect that any reader of this thesis will realize that I have taken this lesson seriously. Jennifer Melcher, my academic advisor, has always expressed her faith in me and my abilities.

I also thank Dianna Sands; her tireless work keeps EPL running. I am extremely grateful for the work of Connie Miller and Leslie Liberman who performed the surgeries for the physiological experiments. Also, I would also like to thank Kelly Brinsko, Craig Atencio, and Brad Cranston for their help with the histological preparations.

I have made so many wonderful friends in Boston that the list becomes too long. These friendships have sustained me. I would especially like to thank all of Bertrand's group: Leonardo Cedolin, Ben Hammond, Ken Hancock, Sridhar Kalluri, Leo Litvak, Martin McKinney, and Zach Smith; these guys provided support that could only be given by people sharing the same experiments and experiences. And then there are the friends who are more like family: Joe Ahadian, Radha Kalluri, Chandran Seshagiri, Irina Sigalovsky, and Jocelyn Songer. These are the people that I relied on, that pushed me

through the really tough times. Domenica Karavitaki (who is the Mom in this big family) scolded me, encouraged me, and cajoled me, whichever I needed at the time, to try harder and to succeed; recently she has even fed me and given me a place to stay. We have experienced some of the happiest moments of our lives side by side. Kostas has patiently put up with the two of us in his kind and generous way.

I thank my real family who has perhaps never really understood what I was doing or why I needed to do it, but supported me anyway. In particular, my mom and dad have always given me the love, encouragement, and support that I needed. My dad taught me integrity, independence, courage, and how to dream big (Texas big). My mom is the source of love and kindness that I depend on through the best and worst times; she is also my best friend. Thanks, Mom and Dad, for always being proud of me. Thanks also to my brother, Clint, for teaching me how to take it as well as I can dish it out. Thanks also to the Rowlands' for accepting me into their family and for giving me Jonathan.

There are no words to express how grateful and thankful I am for the unwavering love and support that Jonathan has given me. He has seen me at my best and my worst and stayed beside me anyway, doing anything that he could to help me through. And he never failed to bring me tea.

Table of Contents

Chapter 1: Introduction.....	7
Background.....	7
<i>Psychophysics.....</i>	8
<i>Cross-correlator model</i>	10
<i>Neurophysiology.....</i>	13
Thesis Overview	16
Chapter 2: The Responses of ITD-Sensitive Units to Signals and Maskers at Different Locations ...	21
Introduction.....	21
Methods.....	22
<i>Recording techniques.....</i>	22
<i>Virtual stimuli synthesis.....</i>	24
<i>Experimental Procedure.....</i>	24
<i>Data Analysis.....</i>	25
<i>Masked Thresholds</i>	27
Results.....	28
<i>A single unit's response to signal and masker placed at different azimuths.....</i>	29
<i>Mean-rate thresholds compared to synchronized-rate thresholds.....</i>	31
<i>The dependence of masked thresholds on signal and masker azimuth</i>	33
<i>Individual unit responses do not show spatial release from masking.....</i>	34
<i>Best thresholds occur for signal at best azimuth</i>	34
<i>Cross-correlator predictions of masking type and signal effect type.....</i>	36
<i>Some, but not all, neural responses match cross-correlator predictions.....</i>	39
<i>Masking type index depends on noise azimuth</i>	40
<i>Signal detection type depends on signal azimuth.....</i>	41
Discussion	41
Chapter 3: A Computational Model of Single Unit Responses.....	63
Introduction.....	63
Model Overview	64
Results.....	65
<i>Rate responses</i>	65
Noise-alone response	65
Signal-plus-noise response for the modified cross-correlator	69
Signal-plus-noise response for the envelope-processor model	70

Envelope-processor model implementation.....	71
<i>Model thresholds</i>	74
Effects of model parameters	78
Discussion.....	82
Chapter 4: Human Behavioral Thresholds Compared to Neural and Model Population Thresholds	108
Introduction.....	108
Methods.....	109
Results.....	109
<i>Human behavioral thresholds and neural population thresholds</i>	109
<i>Model population thresholds</i>	110
<i>CF and best ITD joint distribution affects population thresholds</i>	113
Discussion	116
Chapter 5: Conclusions	126
Thesis Summary.....	126
Validity of Methods	128
<i>The effects of anesthesia</i>	128
<i>Model implementation</i>	130
<i>Definition of single unit and population thresholds</i>	131
Neural Implementation of Mechanisms.....	132
The Central Processor	134
Psychophysics	135
Hearing Aids	136
Automatic speech recognition.....	137
Final Conclusion	137
References	138

Chapter 1: Introduction

Normal-hearing listeners have a remarkable ability to hear in noisy environments. In contrast, hearing-impaired listeners and artificial speech-recognition systems often have difficulty in such environments, making it desirable to find ways to mimic the normal-hearing listener's auditory processing. The ability to hear in noisy environments involves many factors: room reverberation; the listener's attention; familiarity with a speaker's voice, language, and subject; and the number, type, and position of noise sources in the room. For this thesis, we study the underlying neural mechanisms involved in one aspect of noisy-environment listening, termed "spatial release from masking" (SRM), which refers to the observation that a signal is more easily detected when separated in space from a masking noise. To further narrow the problem, we focus on the processing that occurs for SRM at low frequencies (below 2 kHz); these frequencies are important for speech processing and are often spared in listeners with hearing loss. A detailed, quantitative understanding of the neural mechanisms involved in low-frequency SRM may ultimately lead to improved hearing aids, cochlear implants, and artificial speech-recognition systems. We have taken a three-pronged approach in our efforts to understand the neural mechanisms involved in SRM: (1) We recorded from single neurons in the cat inferior colliculus (IC), a convergence center in the auditory midbrain, to determine how these neurons respond to signals and maskers at different spatial locations; (2) we measured human behavioral thresholds for combinations of signals and maskers at different spatial locations using stimuli and methods similar to those used in the physiological experiments; and (3) we created a computational neural model that predicts both the individual neuron responses and the psychophysical performance, providing a link between the physiological and psychophysical results.

Background

In order to frame the research addressed in this thesis, this section describes some of the psychophysical, modeling, and physiological studies that relate to our work. The psychophysical results show that spatial release from masking occurs at all frequencies. At high frequencies, spatial release from masking appears to be achieved primarily through changes in the signal-to-noise ratio (SNR), while the binaural system appears to

improve the signal detection at low frequencies. The ability of the binaural system to improve signal detection at low frequencies has been studied extensively by psychophysicists, using the somewhat unnatural masking-level-difference (MLD) paradigm that involves using in-phase and out-of-phase stimuli. A cross-correlator model developed by Colburn (1973, 1977a, b) describes how the binaural system could achieve the improved detection seen in the MLD studies with a neural network, and a series of studies by Jiang et al. (e.g. Jiang et al., 1997a) shows that low-frequency units in the IC often show responses similar to those predicted by Colburn's model. For this thesis, we explore IC unit responses under more natural conditions, using broadband signals and maskers placed at different spatial locations, and we use the cross-correlator model as a starting point for a model of the neural responses.

Psychophysics

When a sound source is placed in a particular spatial location, the two ears receive slightly different acoustic signals, and these differences give information about the sound's location. For example, interaural time differences (ITDs) are caused by differences in the path lengths from the sound source to the two ears. These path length differences also create interaural phase differences (IPDs) for each frequency component. Interaural level differences (ILDs) are caused by an acoustic head shadow, and filtering by the pinna and the rest of the head and torso can create spectral localization cues. ITDs are the dominant sound localization cue at low frequencies (less than about 2 kHz). At high frequencies, the ITD of the envelope of the signal can also be used for localization. ILDs are larger at higher frequencies (greater than about 2 kHz), while spectral cues, specifically peaks and notches in the head-related transfer functions (HRTFs), are important at even higher frequencies (above about 6 kHz) and are particularly important for changes in elevation (Blauert, 1997).

When a signal and masker are placed at different spatial locations, the listener exploits differences in the stimuli reaching the ears created by the different spatial locations of the signal and masker to improve signal detection, giving SRM. SRM occurs at all stimulus frequencies and may involve different sound localization cues (Bronkhorst and Plomp, 1988; Gilkey and Good, 1995). Because the different sound localization cues

appear to be processed by separate neural pathways, knowing which cues are used to obtain SRM provides information about the underlying neural mechanisms involved. A study by Saberi et al. (1991) showed that listeners are able to improve signal detection with separation of the signal and masker. In this study, the detection thresholds for a broadband signal (a 100-Hz click train) in a broadband masker improved by about 15 dB when the signal was moved away from the masker in space in an anechoic room. For a few conditions, they also compared binaural and monaural thresholds. In these cases, SRM appeared to be due to changes in the signal-to-noise ratio at the better ear. Binaural and monaural thresholds were studied for only a few conditions, however, making it difficult to determine which localization cues lead to improved signal detection for these broadband stimuli.

Gilkey and Good (1995) provide further insight into the cues used to improve signal detection in SRM. Again, they measured SRM for free-field 100-Hz click trains in noise, but in this experiment, the stimuli were filtered to contain low, medium, or high frequencies. They chose their frequency bands “to correspond roughly to frequency regions in which interaural time differences, interaural level differences, and spectral cues, respectively, would be expected to dominate localization performance.” They found SRM of 9 dB for low-, 5 dB for medium-, and 10 dB for high-frequency stimuli when the noise was placed straight ahead (0°); and 8 dB, 12 dB, and 18 dB, respectively, for the noise placed at the left ear (-90°). The signal was moved around in the head in 45° steps, starting at 0° . For low frequencies, ILDs and spectral cues are generally small; therefore, the release seen at low frequencies was most likely due to ITDs. At higher frequencies, however, the release from masking was likely to be due to ILDs and spectral cues. In another study, Good, Gilkey, and Ball (1997) used head-related transfer functions (HRTFs) to measure SRM for similar frequency ranges under both binaural and monaural conditions. Comparing results when the signal and masker were co-located in front to when the signal was moved to the side, the spatial release from masking was similar for monaural (at the better ear) and binaural conditions for both the mid- and high-frequency stimuli, indicating that the signal-to-noise ratio at the better ear largely determines performance in these frequency regions. For low frequencies, however, SRM in the monaural condition was much smaller (4 dB) than in the binaural condition (10

dB). Thus, although binaural sound localization cues did not improve SRM at the higher frequencies, the binaural system seemed to provide a large portion of the SRM observed at low frequencies.

Other sets of psychophysical experiments exploring masking level differences (MLDs) shows that the binaural system can exploit IPDs to improve signal detection. (See Durlach and Colburn, 1978, for a thorough review.) In the most common MLD experiments, identical noise is played to both ears ($N0$); the masked threshold for a signal, usually a tone, played in phase to the two ears ($S0$) is compared to a signal played out of phase to the two ears ($S\pi$). These are the $N0S0$ and $N0S\pi$ conditions, respectively. The difference in masked threshold (the MLD) between these two conditions is about 12-15 dB for a low-frequency (below about 500 Hz) tone in broadband noise. For the case when the noise is anti-phasic instead of the signal (i.e. $N\pi S0$ compared to $N0S0$), there is also a large, although slightly smaller MLD of around 9-10 dB for tones below 500 Hz. Because there is no difference in signal-to-noise ratio between the MLD conditions at either ear, the large improvements can only be due to binaural processing.

Cross-correlator model

The MLD experiments show that the low-frequency binaural system can use differences in IPDs to improve signal detection, but they do not explain the underlying neural mechanisms used to obtain the improvement. Several models have been posed to explain MLDs; one of the first was Durlach's equalization-cancellation (EC) model, which delays and subtracts the inputs from both ears to improve signal-to-noise ratios (see Colburn and Durlach, 1978, for a review). However, Colburn and Durlach show that all of the models they reviewed are essentially mathematically equivalent to Colburn's (1973, 1977a, b) cross-correlator model. The advantage of the cross-correlator model is that it is a physiologically viable mechanism; furthermore, the neurons in the medial superior olive (MSO) appear to be performing the cross-correlations (see below). Colburn's model (Figure 1) is similar to the hypothetical neural network initially proposed by Jeffress (1948). For these models, the stimulus waveform to each ear is filtered by two sets of auditory nerve fibers (ANFs) that span a range of characteristic frequencies (CFs). Pairs of ANFs with the same CFs from each ear provide inputs to an array of delay lines and

coincidence detectors, which fire when the neural inputs from the two sides coincide. The delay lines allow each coincidence detector to respond maximally to a different interaural delay. For example, if a sound has no interaural delay, a unit with equal-length delay lines (a 0-ITD unit) will fire because the inputs to this unit travel the same path length and arrive at the unit at the same time. If, however, the sound arrives at one ear earlier, then a different coincidence detector, one tuned to the ITD of the stimulus, will fire. Effectively, this model performs an instantaneous cross-correlation of the ANF responses.

Using this model, Colburn studied optimal detection strategies for typical MLD conditions and was able to predict most of the known psychophysical results. For example, for the cross-correlator model, the signal detection strategies for the N0S0 condition and the N0S π condition are different (see Figure 2). The example plots show results for a 500-Hz pure-tone signal in broadband noise, filtered through a gammatone filter with a CF of 500 Hz (Johannesma, 1972). The resulting cross-correlations have been scaled to give the percent of the maximal response for the N0 condition. For the N0 condition, the 0-ITD unit will fire maximally because the inputs are exactly in phase; the other units in the network will not fire as much because fewer coincidences occur. In the N0S0 condition (another in-phase condition), the 0-ITD unit fires with an increased rate due to the increased energy; this change in rate is the largest in the population and is therefore likely to dominate the population response.

Adding S π to N0 causes a more complicated change in the neural network's firing pattern. Adding S π causes the overall ITD seen by the unit to fluctuate randomly. To understand how these random fluctuations occur, one can assume that the noise is narrowly bandpass filtered by an ANF with a CF at the tone frequency (Jeffress et al., 1956). With this assumption, the noise is essentially a sinusoid with a frequency equal to the tone frequency, but with an amplitude and phase that varies randomly with time. A snapshot of this sinusoid can be represented vectorially in the complex plane (see Figure 3), although one should imagine the N0 vector changing in length and direction over time. When the signal is added, the filtered version of the stimulus is a sinusoid with randomly varying amplitude and phase due to the noise, but with an additional sinusoid of constant amplitude and linearly-progressing phase due to the signal. For N0S0, the

resultant vectors for the two ears remain in phase regardless of the noise amplitude and phase. For $N0S\pi$ however, the signals are added out of phase to the two ears, producing a randomly varying ITD and ILD between the resultants for the two ears, depending on the instantaneous phase and amplitude of the noise relative to the signal at each ear. Consequently, the firing rate will decrease for the 0-ITD neuron because the signal decorrelates the inputs to this unit (Figure 2). Also, firing increases for the nonzero-ITD units because coincidences occasionally occur at these nonzero-ITDs. Because the most detectable (largest when compared to the variability of the responses) change in correlation occurs when the signal is added out-of-phase, this condition gives the best threshold; the 0-ITD unit's inputs will show the largest change in correlation so this unit is expected to determine the performance of the population.

In contrast to the $N0$ stimulus, the $N\pi$ stimulus will not cause a large response for the 0-ITD unit. Instead, the units with CFs near the tone frequency, F , and best ITDs equal to $1/(2F)$ (1000 μsec for a 500 Hz tone) will produce the largest response (π -ITD units). The maximal response occurs for these units because the response to the signal is strongest and they have best ITDs that counteract the phase inversion of the stimulus at their CF. Because the neuronal delay does not perfectly compensate for the stimulus phase inversion for the broadband noise through the auditory filters, these units' responses to $N\pi$ are a little smaller than those seen in the $N0$ case for the 0-ITD units. When $S0$ is added to $N\pi$, these units show the most detectable change (a decrease) in firing rate, again because the signal further decorrelates the noise response. This change in correlation is smaller than in the $N0S\pi$ condition, resulting in a smaller masking level difference. In summary, the cross-correlator model gives predictions about which units are best for signal detection: essentially, the units yielding the best model thresholds are the ones that response most vigorously to the noise alone and that have the largest change (a decrease) in rate when the signal is added.

In addition to providing insight into signal detection, the Colburn model has also been used to make predictions about speech intelligibility. Because our goal is to improve listening in noisy environments, the ability to link signal detection and speech intelligibility is important. Zurek's (1993) model used predicted improvements in signal detection to predict improvements in speech intelligibility due to separation of the target

and masker. The signal detection predictions were based on differences in the head-related transfer functions (HRTFs) that occur when the signal and masker are at different locations. These HRTFs give the interaural level differences (ILDs) and interaural phase differences (IPDs) for each frequency component arising for sources at different spatial locations. Along with signal-to-noise ratio calculations within frequency bands, Colburn's (1977b) cross-correlator model was used to predict the improvement in signal detection afforded by differences in IPD and ILD from the signal and noise. Then, in a manner similar to that used by Rabiner and Levitt (1967), the predicted signal detection thresholds and the articulation index were combined to successfully predict the overall improvement in speech intelligibility. Therefore, at least for these relatively simple conditions, signal detection can be used to predict speech intelligibility. Overall, understanding if there are neurons behaving like Colburn's cross-correlators for broadband signals and maskers placed at different locations is an important first step towards understanding the underlying neural mechanisms of SRM.

Neurophysiology

As mentioned above, the sound localization cues involved in SRM are thought to be processed by different neural pathways. The divergence of these pathways begins at the level of the cochlear nucleus. Two binaural nuclei in the lower brainstem, the medial superior olive (MSO) and the lateral superior olive (LSO), along with the dorsal cochlear nucleus (DCN) are thought to process the ITD, ILD, and spectral cues, respectively (see below). Because virtually all of the ascending auditory pathways synapse in the tonotopically organized inferior colliculus (IC), the IC represents the first nucleus after the cochlear nucleus that contains nearly all of the information available to the auditory system (see Figure 4 for a diagram of the inputs to the IC). This convergence of information makes the IC an interesting and convenient nucleus for the study of the neural mechanisms of SRM, which occurs over different frequency ranges and uses a variety of different localization cues. The IC is also the first place where the interactions between these input pathways could occur in single neurons. The extent to which the inputs do converge onto individual units in the IC is not known, and it is possible that the inputs are merely relayed to higher centers without interacting (Oliver and Morest, 1984;

Oliver et al., 1997). Regardless of the degree of interaction of the inputs, the responses of the units in the IC almost certainly reflect additional processing beyond that of these lower nuclei, and the nature of this processing is generally not well understood.

As mentioned above, there are (at least) three neural pathways thought to process sound localization cues: the medial superior olive (MSO) is thought to process ITDs, the lateral superior olive (LSO) appears to process ILDs, and the dorsal cochlear nucleus (DCN) has been hypothesized to process monaural spectral cues. The neurons in medial superior olive (MSO) are thought to implement the cross-correlation described by the Jeffress and Colburn models. Units in the MSO receive binaural excitatory inputs and are sensitive to the ITD of tones and noise, and the best interaural phase for MSO units can be predicted from the phases of the monaural responses, as expected for cross-correlator units. As expected for a cross-correlator following a narrow-band filter, these units show a damped-sinusoidal rate response as a function of noise ITD (Goldberg and Brown, 1969; Yin and Chan, 1990). The MSO, a primarily low-frequency nucleus, is probably the dominant input to the low-frequency, ITD-sensitive units that we have studied in the IC. In contrast with the MSO, the LSO is a high-frequency nucleus and is thought to process ILDs through units that are inhibited by the contralateral ear and excited by the ipsilateral ear. These units are thought to “subtract” the contralateral ear’s input from the ipsilateral to emphasize the differences between them, thereby providing ILD sensitivity (Boudreau and Tsuchitani, 1970; Guinan et al., 1972a; Guinan et al., 1972b). Finally, it has been suggested that the contralateral DCN processes the spectral cues in the HRTFs, particularly the notches, using spectral receptive fields with narrow excitatory tuning and inhibitory sidebands (see Young et al, 1992, for a review).

A series of MLD studies (e.g. Jiang et al., 1997a, b; Palmer et al. 1999; Palmer et al., 2000) tested the predictions of the Colburn model for low-frequency units in the anesthetized guinea pig inferior colliculus. Using 500-Hz tones in noise, they began by measuring the responses to the N0S0 and N0S π conditions for both individual units and populations of units (Jiang et al., 1997a, b). They showed that individual units show both positive and negative MLDs, but that when averaged across their sample, the thresholds were better for the N0S π condition compared to the N0S0 condition. As expected from the Colburn model, the units with the best average thresholds have an excitatory response

to the N0 condition and show a decrease in their rate when the anti-phasic signal (i.e., the interaural phase is π) is added. Additionally, the decreases in rate caused by the anti-phasic signal are similar to the changes in response seen for noise with a reduced interaural decorrelation for the majority of neurons (Palmer et al., 1999), validating the overall concept of the cross-correlator model. Finally, for a different unit sample, Palmer et al. (2000) showed that the $N\pi S0$ thresholds averaged across all of their units were better than those seen for N0S0; however, the improvements were less than those seen with N0S π . Overall, these results show that average thresholds across the population for low-frequency units in the IC show that, like the psychophysical results, the N0S0 thresholds are worse than the $N\pi S0$ thresholds, which are worse than the N0S π thresholds. The responses of the low-frequency IC units are generally consistent with the cross-correlator model. However, these authors did report that some units seemed to reflect the effects of additional inhibition; these units had responses that could not be predicted from a simple cross-correlator model.

In summary, the psychophysical results show that SRM occurs at all frequencies, and that the binaural system contributes to SRM for low-frequency stimuli. The extensively studied MLD paradigm shows that the binaural system can improve signal detection when the signal and target differ in IPD. Colburn's cross-correlator model, which uses a system of delay lines and binaural coincidence detectors, predicts most of the known MLD results, giving a hypothetical neural substrate for the processing underlying binaural improvements in signal detection. The physiological results show that, for the MLD paradigm, the cross-correlator model does a good job of predicting the responses of ITD-sensitive units in the inferior colliculus, which presumably reflect coincidence detection occurring in the MSO. However, the MLD condition is somewhat unnatural. Placing a broadband sound source at a given location produces a nearly fixed ITD for all the frequencies; this fixed ITD corresponds to a different IPD for each frequency component. For the anti-phasic condition used in the MLD experiments, the stimulus is out of phase for all frequencies, giving a different ITD for each frequency component. Consequently, it is not entirely clear how the physiological MLD results apply to the more natural situation where broadband signals and maskers are placed at different spatial locations. In this thesis, we study the neural mechanisms of low-

frequency SRM by recording the responses of low-frequency, ITD-sensitive units in the IC to a broadband signal and masker placed at different spatial locations. We also attempt to predict the neural responses with a cross-correlator model. Interestingly, for our experimental paradigm, the cross-correlator model alone fails to predict the results, but a model that includes cross-correlation and additional processing paths can account for the observed results.

Thesis Overview

There are three main chapters of the thesis. Chapter 2 describes the responses of low-frequency, ITD-sensitive units in the cat inferior colliculus to SRM stimuli; Chapter 3 develops a model of these single unit responses based on a cross-correlation model and modulation filters; and Chapter 4 compares the population responses of the single units and the neuron model to human behavioral thresholds. Chapter 5 provides a summary of the results, ties the conclusions of the individual chapters together, and discusses future research directions.

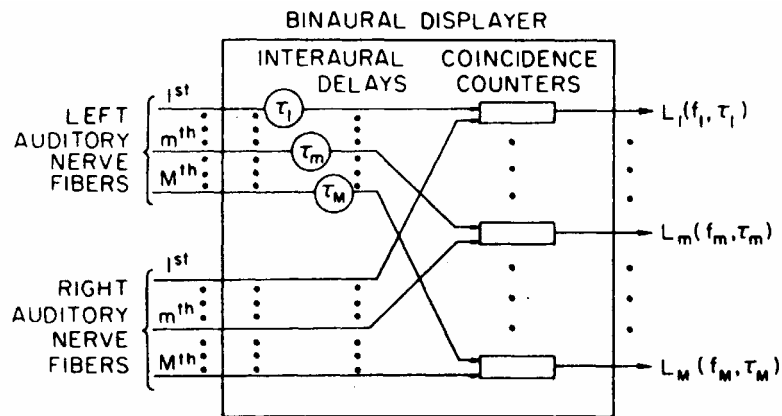


Figure 1: From Colburn (1977a). The cross-correlation model. A left auditory nerve fiber input is delayed by a fixed delay τ_m and then converges on a coincidence detector with a right auditory nerve fiber of the same CF. This combination of delays and coincidence detectors results in a representation of all of the delays and frequencies in a binaural display matrix.

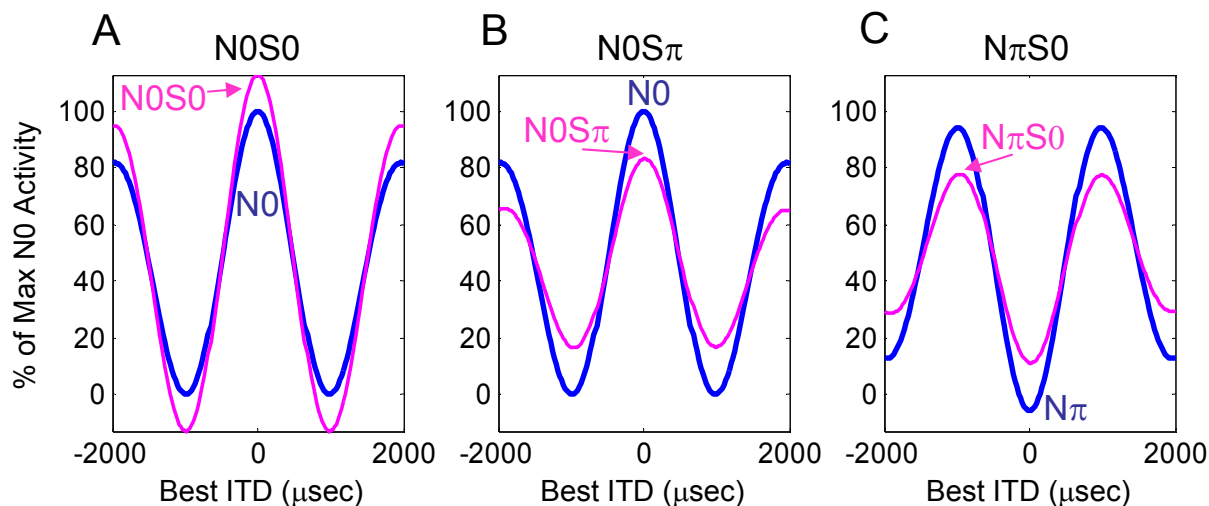


Figure 2: Activity for a set of cross-correlator model units for different masking level difference stimuli. The signal is a 500-Hz tone in noise. We have filtered the stimuli through a gammatone filter with a center frequency of 500 Hz. The x-axis is the best ITD of the model units, and the y-axis is the percent of the maximum activity for the N0 condition. A: N0S0 compared to N0. The addition of the signal causes the largest change in rate, an increase, for the 0-ITD unit (arrow). The additional energy can increase or decrease the firing rate of other units. B: N0S π compared to N0. The largest change in rate again occurs for the 0-ITD unit, but through a decrease in rate. C: N π S0 compared to N π . In this case, the largest change in rate occurs for the π -units, which for a 500-Hz tone, is the one with a best-ITD of 1000 μ sec. Figure adapted from Palmer, Jiang, and McAlpine (2000).

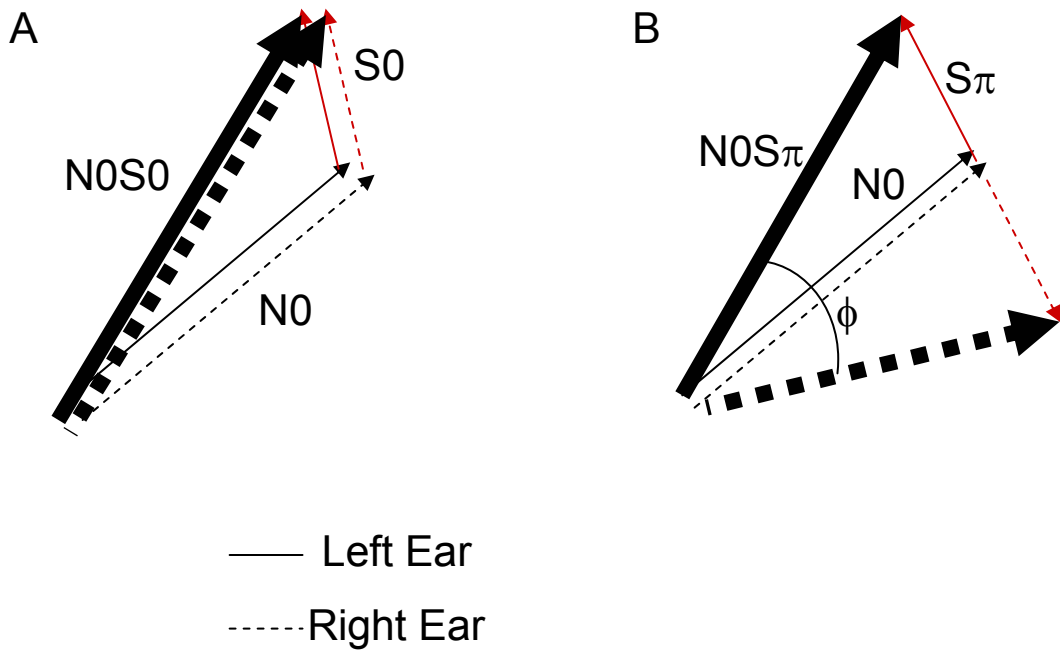


Figure 3: Vector representation of the N0S0 (A) and N0S π (B) conditions. The narrowband noise, N , has an amplitude and phase that varies randomly with time. The tonal signal, S , has an amplitude and phase that is fixed. (A) When S is added in phase to both ears, N0S0, the interaural phase remains zero, regardless of the noise amplitude and phase. (B) When S is added out of phase to the ears, giving N0S π , the interaural phase ϕ varies randomly according to the relationship of S with the random amplitude and phase of N .

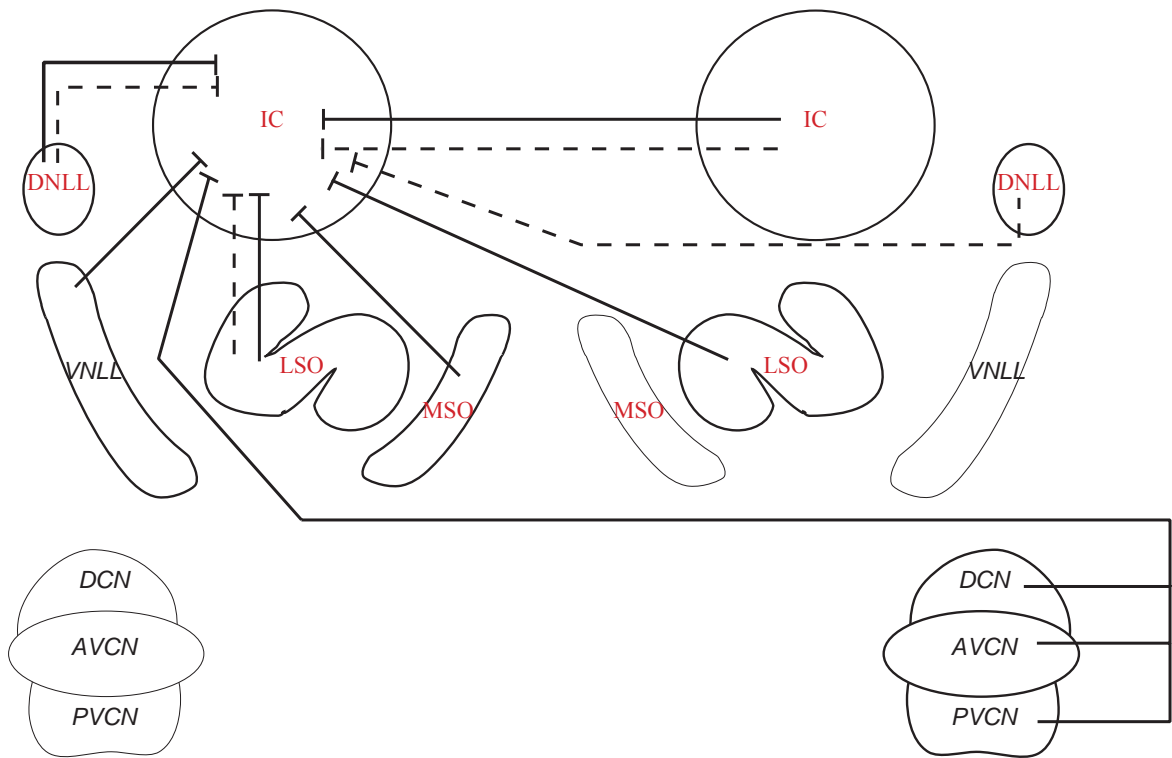


Figure 4: Major projections to the IC. Solid lines indicate excitatory projections, and dashed lines indicate inhibitory projections. Nuclei with labels in red with serifs are primarily binaural; black labels with no serifs are primarily monaural. Note that all of the inputs to these nuclei derive from the auditory nerve; only the connections to the IC are shown, not the other connections between the nuclei. IC: inferior colliculus; DNLL: dorsal nucleus of the lateral lemniscus; VNLL: ventral nucleus of the lateral lemniscus; LSO: lateral superior olive; MSO: medial superior olive; DCN: dorsal cochlear nucleus, AVCN: anteroventral cochlear nucleus; PVCN: posteroventral cochlear nucleus.

Chapter 2: The Responses of ITD-Sensitive Units to Signals and Maskers at Different Locations

Introduction

A listener can more easily detect a signal when it is spatially separated from a masker (Sabeti et al., 1991). This phenomenon, termed “spatial release from masking” (SRM), may contribute to a listener’s ability to hear in noisy environments. Previous psychophysical studies (Good, Gilkey, and Ball, 1997) have shown that binaural hearing contributes to SRM for low-frequency stimuli. The binaural mechanism at these low frequencies is most likely to involve processing of interaural time differences (ITDs). The purpose of this chapter is to explore the underlying neural mechanisms of low-frequency SRM by looking at the responses and masked thresholds of individual ITD-sensitive neurons in the inferior colliculus (IC).

As described in Chapter 1, Jeffress (1948) hypothesized that a set of neuronal delays and binaural coincidence detectors could be the neural mechanism underlying a listener’s sensitivity to ITDs and interaural phase differences (IPDs). Colburn (1977a) showed that such a set of coincidence detectors could be used to exploit differences between tone and noise IPDs to improve signal detection in noise (masking level differences or MLDs), and that a model built from a population of these cross-correlators predicted most of the known psychophysical results. More recently, a series of studies (e.g. Jiang, McAlpine, and Palmer, 1997a, b; Palmer, Jiang, and McAlpine, 2000) showed that ITD-sensitive units in the IC are sensitive to these IPD differences and that in most cases these ITD-sensitive units seem to respond as predicted by Colburn’s coincidence detector model. Specifically, these studies examined the single unit and population responses to stimuli where the signal (in this case, a 500-Hz pure tone) and noise were either in phase (indicated by 0) or out of phase (indicated by π) at the two ears. When compared to the N0S0 (diotic) condition, the single unit thresholds averaged across the population gave a larger release from masking for N0S π than for N π S0, consistent with a large body of psychophysical results (see Durlach and Colburn, 1978, for a review). Furthermore, for the N0S π stimuli, the units with the best thresholds were

those that showed a decrease in their overall rate when the π -phase pure-tone signal was added to the noise, indicating that the signal caused a reduction in the overall interaural correlation seen by these units as Colburn's model predicts. However, these studies did not describe the units' responses to more natural stimuli, such as broadband signals and maskers placed at different spatial locations. Placing a broadband sound-source at a location in space creates a fixed ITD at all stimulus frequencies instead of the fixed IPD used in MLD studies.

In this chapter, we describe how ITD-sensitive units in the IC respond to a 40-Hz chirp-train signal and a broadband-noise masker placed at different (virtual) spatial locations, and we examine the degree to which individual unit responses show spatial release from masking. In contrast to the results shown by Jiang et al. (1997a, b), we show that the best thresholds for chirp trains in noise occur in units for which the addition of the signal increases the overall rate response. We argue that our results do not contradict the results of Jiang et al. (1997a, b), but instead seem to be due to differences between temporal and/or spectral properties of the signals used in the two studies. We also compare how the signal and masker responses change with azimuth to predictions from a cross-correlator model, showing that while the cross-correlator model can predict the neural responses, some important differences arise. Preliminary results from this work have been previously reported (Lane et al., 2003a; Lane et al., 2003b).

Methods

Recording techniques

The responses of single units in the anesthetized cat inferior colliculus were recorded using methods similar to those of Litovsky and Delgutte (2002). Healthy, adult cats were initially anesthetized with an intra-peritoneal injection of Dial-in-urethane (75 mg/kg), and additional doses were provided throughout the experiment to maintain deep anesthesia. Dexamethasone was injected intramuscularly to prevent swelling of the neural tissue. A rectal thermometer was used to monitor the animal's temperature, which was maintained at 37–38°C. A tracheal cannula was inserted, both pinnae were partially dissected away, and the ear canals were cut to allow insertion of acoustic assemblies. A small hole was drilled in each bulla, and a 30-cm plastic tube was inserted and glued in

place to prevent static pressure from building up in the middle ear. The animal was placed in a double-walled, electrically shielded, sound-proof chamber. The posterior surface of the IC was exposed through a posterior fossa craniotomy and aspiration of the overlying cerebellum. Parylene-insulated tungsten stereo microelectrodes (Micro Probe, Potomac, MD) were mounted on a remote-controlled hydraulic microdrive and inserted into the IC. The electrodes were oriented nearly horizontally in a parasagittal plane, approximately parallel to the iso-frequency planes (Merzenich and Reid, 1974). To improve single unit isolation, the difference between the outputs of the two electrodes, which were separated by 125 μm , was often used as the input to the amplifier and spike timer. Spikes from single units were amplified and isolated, and spike times were measured with 1- μs resolution and stored in a computer file for analysis and display.

Histological processing for reconstruction of an electrode track was performed for one cat with a large data yield (see Figure 1). Two out of three 40- μm parasagittal sections of the IC were Nissl-stained, and the remaining were immunostained for calretinin to visualize putative projections from the MSO (Adams, 1995). Staining for calretinin is thought to reveal terminals of MSO axons because the MSO is the only auditory structure projecting to the IC in which calretinin labeling is extensive.

Figure 1 shows three slices: one Nissl-stained slice and the two calretinin slices just medial and lateral to the Nissl slice. In the more medial calretinin slide, there is a small, dark stain, showing the location of the calretinin labeling; a lighter, but larger, but more lightly stained cloud may also be seen (marked by the red oval). The more lateral calretinin slide is stained less darkly, but a faint cloud of stain can still be seen (red oval). The electrode track is evident in the Nissl slice and the more lateral calretinin slice (red arrows), indicating that the track traversed the calretinin region. The microelectrode depths at which we found units (blue arrows) show that we seemed to be recording from the calretinin region, suggesting that these units were getting inputs from the MSO. The other experiments had similar electrode placements and single unit responses; therefore many of the units in our sample are likely to receive MSO inputs.

Virtual stimuli synthesis

Because SRM occurs for stimuli of all frequency ranges (Gilkey and Good, 1995), we use head-related transfer functions (HRTFs) to simulate sounds at different azimuths. This chapter (and the remaining ones) focuses on low-frequency neurons that are sensitive to ITD, the primary sound localization cue at low frequencies. The HRTFs represent the directionally dependent transformations of sound pressure from a specific location in free field to the ear canal. Virtual-space stimuli were synthesized in the same manner as Litovsky and Delgutte (2002) using HRTFs measured in one cat by Musicant et al. (1990) for frequencies above 2 kHz and a spherical-head model for frequencies below 2 kHz. Specifically, the low-frequency HRTFs were the product of two components: 1) a directional component representing acoustic scattering by the cat head was provided by a rigid-sphere model (Morse and Ingard, 1968, p. 418–422); and 2) a non-directional, frequency-dependent gain representing the sound pressure amplification by the external ear was derived from measurements of acoustic impedance in the cat ear canal (Rosowski, Carney, and Peake, 1988). Using a frequency-dependent weighting function, the model HRTF for frequencies below 2 kHz was joined with the measured HRTF above 2 kHz to obtain an HRTF covering the 0- to 40-kHz range.

The signal used was a 40-Hz, 200-msec chirp train presented in continuous noise (see Figure 2). Each chirp's frequency was swept from 300 to 30 kHz logarithmically and had an exponentially increasing envelope designed to produce a flat power spectrum; both signal and noise contained energy from 300 Hz to 30 kHz. In some cases, we also used 100-Hz click trains as signals more like the stimuli used in the psychophysical literature (Saber et al., 1991, and Gilkey and Good, 1995); however, units in the IC often responded with higher and more sustained rates to the chirp trains, presumably due to the lower repetition rate of the chirp train. Only results from the 40-Hz chirp trains are presented here.

Experimental Procedure

Search stimuli were usually 200-msec, 43 dB SPL chirp-trains with a repetition rate of 2 Hz; however, 200-msec broadband noise bursts, also with a repetition rate of 2 Hz, were used in some of the earlier experiments. Both the azimuth and the laterality (binaural or

monaural) of the search stimulus were varied during the experiments in an effort to find a larger number of units and a more varied sample. Once a unit was located, a frequency tuning curve was measured by an automatic tracking procedure (Kiang and Moxon, 1974) to determine each unit's characteristic frequency (CF_{TC}).

A noise-delay function was also measured: the rate was measured as a function of the ITD of one burst of “frozen” noise (Figure 3, solid line with error bars). The ITD was usually varied from -2000 μ sec to 2000 μ sec with a step-size of 400 μ sec, although we often sampled the ITDs inside the physiological range (-290 to 290 μ sec) more finely. For the noise-delay function, the noise bursts were usually 200 msec in duration, although shorter durations were used in some experiments.

The primary measurements in these experiments were the signal-plus-noise and noise-alone responses as a function of noise level (see Figure 4), which we use to compute the single-unit masked thresholds. In order to see the effect of the masker on the signal response, the signal level was fixed near 40 dB SPL, and the noise level was varied in 6 dB steps in randomized order for both a 200-ms signal-plus-noise condition and an immediately following 200-ms noise-alone condition. A different noise sample was used on each trial, but the same set of samples was used for all noise levels. The signal was usually repeated 16 times for each noise level; the signal repetition rate was 2.5 Hz. The response as a function of noise level was measured for several noise azimuths, as time permitted, again presented in a randomized order. The signal azimuth was initially fixed at a location with a strong excitatory response, usually on the side contralateral to the recording site (positive azimuths). The responses for other signal azimuths, including some at unfavorable azimuths, were measured if time permitted. The signal azimuths used were -90°, 0°, 45°, and 90°, and the noise azimuths used were -90°, -54°, -45°, 0°, 18°, 36°, 45°, 54°, 72°, and 90°.

Data Analysis

A unit was included in this study if it was low-frequency ($CF_{TC} \leq 2.5$ kHz), gave a sustained response to chirp trains at some signal azimuth, and was sensitive to ITD. We considered a unit ITD-sensitive if the noise-delay function was modulated by at least 50%, i.e., if the minimum rate was less than half of the maximum rate.

To determine the best ITD and characteristic frequency (BF_{ITD}) for each unit, we fit the noise-delay function with a Gabor function (McAlpine and Palmer, 2002), which is an exponentially damped cosine:

$$G = Ae^{\frac{-(ITD-BITD)^2}{2s^2}} \cos(2\pi * BF_{ITD}(ITD - BITD)) + B$$

The least-squares fit was obtained using Matlab's *leastsq* function. We constrained the best ITD for the fine structure (inside cosine) to be equal to the best ITD for the envelope (in exponential) to facilitate the modeling studies described in subsequent papers. The Gabor fit gives estimates of the characteristic frequency (BF_{ITD} in the equation above) and the best ITD ($BITD$) of the unit. The additional parameters, A , B , and s , determine the amplitude of the curve, the offset of the curve, and how quickly the damped cosine decays, respectively. Additionally, the Gabor function was half-wave rectified to ensure that the rate is never negative. A noise-delay function and its Gabor fit are shown in Figure 3. We show the Gabor function without half-wave rectification to emphasize the location of the worst ITD, which is not obvious from the rate function itself. In a few cases (4 out of 31), because the noise-delay functions were not always sampled as finely as the rate-azimuth functions, the Gabor fit gave results that did not give the best and worst azimuths at the correct locations. In these cases, the best ITD and BF_{ITD} were fit by hand to give appropriate best and worst ITDs to match the best and worst azimuths. In some experiments, the left noise burst was shorter than the right due to a programming error. In this case, the rate window for the ITD-sensitivity analysis was adjusted to match the length of the shorter noise burst to the two ears. A Gabor function could still be fit in all but one case. For that one unit, the ITD function, although fully modulated, was so noisy that the Gabor fitting procedure did not converge after 1000 iterations; this unit was not included in the sample response. It is possible that some other ITD-sensitive units were excluded from our sample based on their seeming lack of ITD sensitivity due to this programming error.

The upper axis in Figure 3 shows the noise ITD while the lower axis in Figure 3 is in units of relative IPD (ϕ) in cycles, which is defined by

$$\phi = (ITD - BITD) * BF_{ITD}$$

By normalizing the axes by both the best ITD and the BF_{ITD} , we have dimensionless metric where the unit's preferred phase is 0 cycles, corresponding to the unit's best ITD, and the unit's worst phase is -0.5 cycles, corresponding to the unit's worst ITD given by

$$WITD = BITD - 1/(2 * BF_{ITD}).$$

We use cycles for relative IPD so as not to confuse the reader with azimuth, which is expressed in degrees. Using relative IPD is convenient for comparing the responses of units with different CFs and best ITDs.

Masked Thresholds

The masked threshold was defined as the signal-to-noise ratio (SNR) at which the signal can be detected for 75% of the stimulus repetitions. As shown in the dot rasters in Figure 4, the second column, the addition of the signal can be detected in many ways, depending on the signal and masker levels and configurations: for example, the signal can cause an increase (e.g. Column A) or decrease (e.g. Column C) in rate in the signal-plus-noise window, a change in the spike arrival times in the signal-plus-noise window (e.g. Columns A), and/or a suppression of the spikes in the earliest part of the noise-alone window (e.g. Column A for the noise at 37 dB SPL). It is possible that a central processor could use any or all of these cues to detect the signal, and, in the best case, the central processor would be able to use the best combination of cues, perhaps though the use of a signal template for each condition, to detect the signal in each case. Given that we only had a few stimulus presentations in each condition, developing an accurate signal template was difficult; instead, we chose to detect the signal through more traditional methods involving changes in mean rate and spike times.

Two different response metrics were used to detect the signal: mean rate and synchronized rate. Mean rate is simply the number of spikes in the measurement window, and the synchronized rate is the Fourier component of the response at the signal repetition rate, 40 Hz (Kim and Molnar, 1979). The synchronized rate, which is also the mean rate multiplied by the vector strength as defined by Goldberg and Brown (1969), includes information about the spike timing as well as the overall rate. Figure 4 shows the mean rate (third row) and the synchronized rate (fourth row) as a function of noise level for one unit (see results for more details). To determine the masked thresholds, we

calculated the percent of stimulus presentations for which the mean rate/synchronized rate was greater in the signal-plus-noise window compared to the noise-alone window (Figure 4, bottom row). The percent curves were converted to their equivalent z-score, smoothed with a three-point triangular filter, and then converted back to percent. Thresholds (circles in Figure 4, bottom row) could occur at either 75% or 25% (dashed lines) because a signal could be detected through either an increase or decrease in rate. As we describe more fully in the Results section, for the cases when the signal and masker were separated, the mean rate thresholds yielded better or equal thresholds than the synchronized rate thresholds. Therefore, unless explicitly stated otherwise, masked threshold refers to the mean rate threshold.

We determine confidence intervals for the masked thresholds using bootstrapping methods. For each noise level, we sample with replacement the spike trains for the different stimulus presentations, obtaining a bootstrapped set of stimulus presentations. We then recompute the percent curves for the new, bootstrapped set of spike trains and recalculate the thresholds. The threshold is recomputed in this way 100 times. The error bars for the masked thresholds are then the thresholds from the 10th to the 90th percentile. Additionally, we required that, for a 25% threshold to be used, 80 out of 100 of the bootstrapped thresholds had to occur because the signal decreased the overall rate. If this was not the case, then the 75% threshold was used. This requirement eliminated very high thresholds that occurred due to spurious estimates of percent-correct points. For the final threshold estimate, we took the median for all the bootstrapped percent curves as the final percent correct curve and measured the threshold for this median curve.

Results

We recorded masked threshold curves for 45 units in the inferior colliculus. After eliminating the units that were not ITD-sensitive and/or had CF_{TCS} above 2 kHz, there remained 31 units from 10 animals with characteristic frequencies (CF_{TCS}) between 250 and 1300 Hz. Figure 5 shows the best and worst ITDs for the units in our sample as a function of BF_{ITD} (as determined through with the Gabor fits). As found by McAlpine, Jiang, and Palmer (2001) in guinea pigs, best ITDs tend to decrease with increasing BF_{ITDS} , and most of the best ITDs (squares) and worst ITDs (x's) are outside the

physiological range (dotted lines). The physiological range here is defined as $\pm 290 \mu\text{sec}$, which are the ITDs for $\pm 90^\circ$ in our HRTFs.

A single unit's response to signal and masker placed at different azimuths

Before showing how the responses and thresholds vary with azimuth, we first show that these units' sensitivity to azimuth is due primarily to their sensitivity to ITD. Figure 6 (left) shows that one ITD-sensitive unit's rate response is similar for noise varying in azimuth (black solid curve) and ITD (blue dash-dot curve), although the rates for the ITD function were higher than those of the rate-azimuth function in some cases. This unit shows a maximum firing rate for the noise at $+90^\circ$ (contralateral to the recording site), which is the hemifield preferred by most IC units. For each azimuth, the ITD was chosen to match the delay yielding the maximum of the interaural cross-correlation in the HRTFs. The right panel shows the rate when the ITD alone was varied compared to the rate when the azimuth was varied for 19 of the units in our sample. The dashed line shows where the two rates would be equal. In general, the two responses are similar, indicating that ITD largely determines these units' azimuth sensitivity.

In addition to the noise-alone response changing with azimuth, the noise's effect on the signal response and the signal's effect on the noise response also change with azimuth. Figure 4 shows the signal-plus-noise response and the noise-alone response for three signal and masker configurations for unit 22-0, which had a BF_{ITD} of 740 Hz and best ITD of $290 \mu\text{sec}$. As expected given the unit's best and worst ITDs (which were either at or beyond physiological range of the cat), the unit's best azimuth was $+90^\circ$, and its worst azimuth was -90° . The first row shows the three signal and masker configurations: the signal and masker co-located at $+90^\circ$ (Column A, S+90, N+90); the signal at $+90^\circ$ and the noise at -90° (Column B, S+90, N-90); and the signal at -90° and the noise at $+90^\circ$ (Column C, S-90, N+90).

The second row shows the temporal discharge patterns for the unit as a function of noise level for each of these signal and masker configurations. In these dot rasters, every dot represents a spike, and the stimulus presentations for different noise levels are separated by the solid lines. At low noise levels (bottom of the rasters) and for the signal at $+90^\circ$ (Figure 4, Columns A and B), the unit produces a synchronized response to the

40-Hz chirp train. For the signal at -90° , however, the response to the signal is much weaker, consisting of only an onset response at the lowest levels. As the level of the noise is raised, the signal response is either overwhelmed (excitatory or “line-busy” masking, Columns A and C) or suppressed (suppressive masking, Column B) by the noise response.

So that we can compare the thresholds for different response patterns, we define two quantities which describe how the noise masks the signal response and how the signal is detected. The mean rates for the signal-and-noise interval and the noise-alone interval for these signal and masker conditions are shown in the third row. The “masking type index” quantifies the effect of the noise masker at threshold. The MTI is the difference between the signal-in-noise rate at threshold, $R(S+N_{Th})$, and the approximate signal-alone rate, $R(S)$, the signal response with the noise at the lowest level, which is then normalized by whichever of the two rates is larger:

$$MTI = \frac{R(S + N_{Th}) - R(S)}{\max(R(S + N_{Th}), R(S))}$$

The MTI ranges from -1 to 1 . For Column C, the noise overwhelms the signal response, the difference in rates is positive, and the MTI of 0.79 is near $+1$, indicating excitatory masking. For Column B, the masker suppresses the signal response without exciting the unit itself, and the MTI of -0.86 is near -1 , indicating suppressive masking. For Column A, however, although both the signal and masker are at the best azimuth, the masker at first decreases the signal response, but then eventually overwhelms the signal response; in this case, the MTI is negative, but smaller in magnitude than Column C (-0.33).

The rate functions also show that the addition of the signal can increase (Columns A and B) or decrease (Column C) the overall rate, depending on the signal azimuth. We defined an index to quantify the effect of the signal on the noise response, the signal effect index (SEI). This index is again a normalized difference, this time between the signal-plus-noise rate, $R(S+N_{Max})$, and the noise-alone rate, $R(N_{Max})$, at the noise level where the signal causes the largest change in rate:

$$SDI = \frac{\max(R(S + N) - R(N))}{\max(R(S + N_{Max}), R(N_{Max}))}$$

To obtain the SEI, we looked only at changes in rate that had the same sign as the change in rate caused by the signal at threshold. For Columns A and B, the signal has the largest change in rate, a positive one, at the lowest noise levels, giving SEIs of 1 in both cases, but for Column C, the largest (negative) change occurs for noise levels around 45 dB SPL, giving an SEI of -0.63. The magnitude of the SEI in Column C is less than 1 because the signal does not completely suppress the noise response.

The fourth row of Figure 4 shows the synchronized rate for the signal-plus-noise response as well as the noise-alone response as a function of the noise level. Because the synchronized rate is the mean rate multiplied by the magnitude of vector strength at 40-Hz, the synchronized rate is the same as the mean rate when the response is perfectly synchronized to the signal (all three conditions at low levels, Column B at all levels). Otherwise, the synchronized rate is less than the mean rate, an effect that can make the thresholds better or worse, as shown below. Finally, row 5 shows the synchronized rate (red x's) and the percent-greater functions for the mean rate (black dots) as a function of noise level; percent-greater is the percent of the stimulus presentations for which the mean rate/synchronized rate is larger in the signal-plus-noise window than in the noise-alone window. Threshold is defined as the noise level where the signal can be detected for 75% of the stimulus presentations through either an increase or decrease in mean rate/synchronized rate, which is noted by the dotted lines at 75% and 25%. The circles show the thresholds for each of the conditions (see Methods for more details). Notice that the synchronized rate thresholds can be higher (Column A), the same as (Column B), or lower than (Column C) the mean rate thresholds.

Mean-rate thresholds compared to synchronized-rate thresholds

We used two methods for measuring masked threshold: one based on mean rate and one based on synchronized rate. When assessed in this way, spike timing can, but does not always, improve the masked thresholds, as shown in Figure 4. When both the signal and the noise are excitatory (Figure 4, Column A), the use of timing information can improve thresholds because, at threshold, the spike times may become synchronized to the signal modulation without changing the mean rate compared to the noise-alone condition. When the noise suppresses the signal response (Figure 4, Column B), the percent greater

curves and the masked thresholds for the mean rate and synchronized rate are similar because the timing provides no additional information. Finally, when the noise is excitatory and the signal is suppressive (Figure 4, Column C), the synchronized rate threshold is actually worse than the rate threshold because the signal acts to decrease the rate but increase the vector strength, reducing the overall difference between the signal-plus-noise response and the noise-alone response. Consequently, the percent-greater curve did not reach 25% so the 75% threshold had to be used. It is possible that using some other method of incorporating spike times would be more effective in this condition.

Figure 7, Panels A and B, shows both the synchronized-rate thresholds and the mean-rate thresholds as a function of noise azimuth for S+90 and S-90 for unit 22-0 (Figure 4). Again the synchronized-rate thresholds can be the same, better, or worse than the mean-rate thresholds, depending on whether the signal and masker are excitatory or suppressive. Panels C and D show how the differences in synchronized-rate and mean-rate thresholds are related to the MTI for the entire sample. Panel C shows the difference in thresholds when the signal is detected through an increase in rate (positive SEIs) as a function of MTI. For suppressive masking (negative MTIs), there is little difference between the two thresholds; for excitatory masking (positive MTIs), the synchronized-rate thresholds are better. In contrast, when the signal is detected through a decrease in rate (negative SEIs, Panel D), the synchronized-rate thresholds are usually worse than the mean-rate thresholds. In this case, the masking cannot be suppressive as there would be no neural response to the signal or the noise.

In Chapter 4, we compute a population response based on the threshold for the most sensitive unit across the population. As we show below, the best overall thresholds occur when the signal is at a favorable azimuth and the noise suppresses the signal response, the case when the rate and synchronized rate thresholds are the same and as a result give very similar population results. To compare our results to the previous physiological and psychophysical results, we would like to compare these thresholds to the cases where the signal is at unfavorable azimuths, where the synchronized-rate thresholds are worse than the mean-rate thresholds. The cases where the synchronized-rate thresholds are best, when both the signal and masker are excitatory, are the worst

thresholds overall. We will therefore use the mean-rate thresholds from this point on because they give better or equal threshold estimates for the conditions that yield the best thresholds and allow for a more fair comparison between the thresholds for the signal at the best azimuth and the signal at the worst azimuth. Because of our decision to ignore the spike timing information, the change in thresholds between the best conditions (where using spike times would not improve thresholds) and the worst conditions (where spike times would improve thresholds) is larger for the mean rate thresholds than the synch rate thresholds. As a result, our estimates of the amount of SRM for single units may be somewhat overestimated if the signal is actually detected based on a change in synchronized rate.

The dependence of masked thresholds on signal and masker azimuth

Masked thresholds for unit 22-0 (Figure 4) are shown in Figure 8, top panel, as a function of noise azimuth for four signal azimuths. For the signal at 45° and 90° , moving the noise away from the signal can improve thresholds by up to 20 dB. However, when the signal is at 0° , thresholds become slightly worse as the noise moves from the midline to the contralateral (positive azimuth) side. For the signal at -90° , the thresholds increase and then decrease as the noise is moved away from the signal. This change occurs because the signal is detected through a decrease in rate for the noise azimuths greater than zero, but through an increase in rate for the noise at 0° and -45° . For the noise at 0° , the signal transitions between causing a decrease in the rate for positive noise azimuths to an increase in rate for negative noise azimuths. As a result, the signal's effect on the noise-alone response for this condition is weak, making this threshold the worst of all the conditions. In this case, units in the other IC would be expected to have good thresholds.

This unit's response shows that, by separating the masker from the signal, the masked thresholds can be changed by 20 dB, although this improvement does not occur for all signal azimuths. A histogram of the masked threshold changes that occur when the masker is separated from the signal for all the units in the sample is shown in the bottom panel of Figure 8. We do see improvement in thresholds in many, but not all, cases. As shown in the next section, the observed improvement does not seem to be

related to the separation of the signal and masker directly, but instead seems to be a consequence of placing the signal and masker at favorable and unfavorable azimuths.

Individual unit responses do not show spatial release from masking

Separation of the signal and masker does not necessarily improve an individual unit's masked thresholds. If separation improved an individual unit's thresholds, then the noise azimuth with the worst threshold would be equal to the signal azimuth. Instead, the worst thresholds tend to occur when the noise is at the best noise azimuth, the one that yields the most excitation. Figure 9 shows the noise azimuth with the worst threshold, the "worst-threshold noise azimuth," as a function of the signal azimuth (left panel) and the best azimuth (right panel, defined as the azimuth with the relative IPD nearest to 0). The correlation between a unit's worst-threshold noise azimuth and the signal azimuth is 0.15, which is not significantly different from 0 as determined in a two-sided t-test ($p > 0.01$). Therefore, the worst thresholds do not necessarily occur when the signal and masker are co-located. In contrast, the correlation between the worst-threshold azimuth and the best noise azimuth is 0.57, which *is* significantly different from 0 ($p < 0.001$), indicating that strong excitation by the masker tends to produce poor masked thresholds. Overall, the individual unit responses do not show a correlate of spatial release from masking. However, because the units show a variety of rate-azimuth functions, a correlate of spatial release from masking may exist in the response of a population of these neurons (see Chapter 4).

Best thresholds occur for signal at best azimuth

For most of the units in our sample, $+90^\circ$ is near the units' best ITD, and -90° is near the units' worst ITD. Placing the signal and masker at these azimuths is therefore somewhat analogous to the well-studied $N0S0$, $N0S\pi$ and $N\pi S0$ conditions: the in-phase conditions ($N0$, $S0$) are similar to placing the stimulus at the best azimuth because the stimulus would appear in phase to the neuron, and the out-of-phase conditions ($N\pi$, $S\pi$) are similar to placing the stimulus at the worst azimuth because the stimulus would appear out-of-phase to the neuron. The psychophysical thresholds for the $N0S\pi$ condition are better than the $N\pi S0$ for a wide variety of signals, and both thresholds are better than the

thresholds for the NOS0 condition. Jiang et al. showed a correlate of this threshold hierarchy with their unit populations for a 500-Hz pure-tone signal. Furthermore, they showed that the individual unit thresholds were generally better when the signal decreased the overall response, as expected in the NOS π condition, presumably because the change in interaural correlation was the most detectable in this case. (See Introduction.) If these results can be extended to our units and stimuli, one might expect that the best threshold would occur when the signal is placed at the *worst* azimuth, -90° , and the noise was placed at the *best* azimuth, $+90^\circ$. (It is important to realize that if, for a given unit, the best ITD lies outside the physiological range, then $+90^\circ$ will be the best azimuth. However, placing the stimulus at this azimuth does not necessarily result in a perfectly correlated input to this unit.) Figure 10 shows the thresholds for 11 units in three animals for which we measured thresholds for the signal and noise on opposite sides of the head: in one case, the signal was near the best azimuth (S+90, N-90, white squares), and in the other, the signal was near the worst azimuth (S-90, N+90, black circles). We also show the threshold when the signal and masker are co-located at $+90^\circ$ (S90, N90, blue diamonds). The abscissa is the difference in the absolute value of the relative IPD for stimuli placed at -90° and $+90^\circ$, showing the amount of change in IPD at the unit's BF_{ITD} between the two positions. Note that none of responses are completely out of phase (0.5), which would require the best ITD to be exactly that for $+90^\circ$ and the worst ITD to be exactly that for -90° . However, for all of the units, the thresholds for the signal placed near the *best* azimuth (Figure 10, squares), the condition most like N π S0, are about the same or better than the thresholds for the signal placed near the *worst* azimuth, the condition most analogous to NOS π , the reverse of the relationship seen in the MLD psychophysical and physiological studies. The S+90, N-90 thresholds are always better than the co-located thresholds as expected from the MLD psychophysical and physiological results; however, the S-90, N+90 thresholds, which might be expected to be the best thresholds overall, are not necessarily even as good as the co-located thresholds.

To better understand the difference between these results and those of Jiang et al. for pure tone stimuli, we also measured the masked thresholds for a 500-Hz pure tone for one unit (BF_{ITD} of 600 Hz, best ITD of 400 μ sec) and compared them to the chirp-train thresholds. The dot rasters for this unit (26-5) show that the overall rates for the pure

tone signal (Figure 11, Panel E) were higher than the responses to the chirp-train (Panel A) when the signals were placed at favorable azimuths. The pure tone signal was also more effective at suppressing the noise response (Panel F) than the chirp-train signal (Panel B) when the signals were placed at an unfavorable azimuth. When the tone was used as a signal for this unit, the threshold SNR for the signal at -90° was *better* than when the tone was placed at $+90^\circ$ (Figure 12), consistent with the Jiang et al., (1997 a, b) results. Therefore, our results do not contradict the previous results using pure-tone signals. Rather, the differences between the signals, presumably differences in either the temporal envelope or spectral composition, seem to change the threshold hierarchy.

Because there are also differences between the pure-tone and the chirp-train signals in the amount of signal energy that passes through the auditory filters, we varied the level of the chirp train for this same unit to see if the signal level affected the thresholds. As the level of the chirp train was raised to 58 dB SPL, the threshold SNR for the signal at $+90^\circ$ remained about the same, but the threshold SNR for the signal at -90° improved (dot rasters, Figure 11, Panels C and D; thresholds, Figure 12). For a low-level signal at -90° , the signal does not suppress the noise response enough to be detected, but at the higher signal levels, the amount of suppression increases, and the thresholds are about the same as the case when the signal was at $+90^\circ$. Thus, the differences in results for the tone and chirp train signals do not seem to be entirely caused by the differences in overall signal energy through the auditory filters; instead, other differences in the spectral or temporal properties seem to be affecting the outcome.

Cross-correlator predictions of masking type and signal effect type

In order to determine how the neural responses compared to those of the well-studied cross-correlator model proposed by Jeffress (1948) and Colburn (1977a), we implement the cross-correlator shown in Figure 13. Our goal here is not to implement a detailed model that produces accurate thresholds (for a more complete model, see Chapter 3), but to simply see how the cross-correlator model's responses changes with signal and masker azimuth. In the next sections, we compare these predictions to the actual neural responses.

For the model, we first pass the stimulus waveform to both ears through an auditory nerve fiber (ANF) model. We use the Zhang et al. (2001) model, which gives appropriate results for broadband stimuli, for an ANF with a 50 spikes/sec spontaneous rate. We then delay the output of one ear and then multiply the two signals and integrate them over time to get the correlation evaluated at the best ITD. Because the output of the ANF model is a probability of discharge and because we do not subtract the mean before computing the correlation, the correlation is always positive. Since it is unnormalized, the correlation depends on the overall level. For these predictions, we assume that a cross-correlator unit's rate response is monotonically related to the unnormalized correlation. Below the model diagram, we show the damped-sinusoidal correlation as a function of noise ITD for one model unit.

For this simulation, we used a BF_{ITD} of 740 Hz and a best ITD of 290 μ sec to match the response of unit 22-0 (the unit in Figure 4). These parameters make $+90^\circ$ the best azimuth, and -90° the worst azimuth. The output cross-correlation for the signal and masker placed at -90 , 0 , and $+90^\circ$ in all combinations are shown in Figure 14; the rows have the same signal azimuth, and the columns have the same noise azimuth. The correlation values are averaged over 16 different noise samples. We defined threshold for this unit's simulations to be the signal-to-noise ratio where the difference in correlation between the signal-plus-noise condition and the noise-alone condition was 0.05, a value chosen to give threshold SNRs similar to those obtained from an actual unit with this BF_{ITD} and best ITD.

The noise-alone response (blue dash-dot lines) generally increases with level due to the lack of normalization. The correlation changes predictably with noise azimuth: for N-90, the worst azimuth, the correlation remains near 0; for N0, the correlation rises to an intermediate value; and for N+90, the best azimuth, the correlation reaches the maximum value seen.

The signal-plus-noise response also changes with noise level. For low noise levels, the response is dominated by signal-alone response: for S-90, there is little response; for S0, the response is moderate; and for S+90, the response is large. As the noise level increases, the signal-plus-noise response gradually changes from the signal-alone response to the noise-alone response. When the signal and masker are co-located,

the signal can only be detected through an increase in overall energy (plots on the diagonal starting in the upper left corner). For S-90, N-90, both the signal and the masker have near-zero correlation so that the signal cannot be detected at any noise level. For the other two co-located conditions, the signal is detected through an increase in the correlation due to the increase in energy. In these cases, the masker eventually swamps the signal response, giving an example of excitatory masking. For S+90, N+90, the signal-plus-noise correlation decreases slightly before rising again as the noise level is raised, due to adaptation in the ANF model responses. Nevertheless, for both of these co-located conditions, the MTI was positive, indicating excitatory masking. The SEI was also positive because the signal was detected through an increase in rate.

When the signal was at a more favorable azimuth than the noise (below the diagonal), the signal was again detected through an increase in rate, giving a positive SEI. The MTI is negative because the noise decorrelates the signal response, bringing the overall rate down; for N-90, the noise masks the signal response without producing much of an excitatory response of its own, giving a negative MTI.

When the noise was at a more favorable azimuth than the signal (above the diagonal), the results were more varied. For S-90, N+90, the signal is detected through a decrease in rate because the signal decorrelates the noise response, making the SEI negative; the magnitude of the SEI is small because the signal produces only a small change in the rate. The MTI is positive because the rate at threshold is higher than the signal-alone response. For S-90, N0, the signal does not decorrelate the noise response enough to be detected at any noise level. For S0, N+90, because the signal alone produces a response, the signal is detected through an increase in rate, although the noise is at a more favorable azimuth than the signal. Again the MTI is positive because the noise swamps the signal response.

Figure 15 shows the cross-correlator's predictions for another unit, this one with a BF_{ITD} of 500 Hz and a best ITD of 925 μ sec. These parameters make +90° the best azimuth, as before, but the worst azimuth is now -22°; -90° becomes a favorable azimuth. The predictions reflect the differences: the responses for both the signal and noise at 0° are small and are larger at +90 and -90°. In this case, the threshold criterion was set to 0.01; again this value was chosen to give SNRs at threshold similar to those seen for the

unit with the same best ITD and BF_{ITD} . The MTI (when it could be measured) is positive for the favorable noise azimuths ($+90^\circ$ and -90° , first and third column) and negative for the unfavorable noise azimuth (0° , second column). In contrast to the previous set of predictions, for these parameters, there is no condition where the signal is detected through a decrease in rate; for S0, N90, where the signal is at an unfavorable azimuth and the noise is at a favorable one, the signal appears to decrease the overall rate somewhat, but not enough to reach the threshold criterion.

Overall, the SEI and the MTI for the cross-correlator model are generally predictable from the locations of the signal and masker compared to the individual unit's best and worst azimuths. In particular, for the signal at a favorable azimuth, the masking is usually suppressive for unfavorable azimuths (the MTI is negative) and excitatory for favorable azimuths (the MTI is positive). Similarly, for favorable noise azimuths, the signal decreases the overall rate for unfavorable signal azimuths (the SEI is negative) and increases the overall rate for favorable signal azimuths (the SEI is positive). We now compare these predictions to the neural responses, first for two individual units and then for the entire sample.

Some, but not all, neural responses match cross-correlator predictions

Figures 16 and 17 show the actual responses of the two units whose response we predicted using the cross-correlator model. The response of unit 22-0 in Figure 16 is similar to the prediction: the noise response increases with noise level and azimuth, the signal response increases with signal azimuth, and the masking types and the signal effects on the masker are generally correct. There are three main differences between the actual responses and the predictions: 1) for S-90, the signal suppresses the response more than predicted, 2) for S-90, N0, even though the signal is near the worst azimuth, the signal produces an excitatory response at low levels that allows the signal to be detected through an increase in rate, and 3) for favorable signal and masker azimuths, the decrease in signal-plus-noise rate with increasing noise levels was larger than predicted, indicating that the neural adaptation may be greater in the neural response than in the model.

The predictions for 22-26 (Figure 17) are less good. Unfortunately, the response to S-90 was not measured for this unit so we cannot compare that case. For the other

conditions, although the cross-correlator predicts that the noise-alone response should be larger than the signal response at high noise levels, the actual response remains non-monotonic and remains near 0. The response to the signal-alone changes with signal azimuth, from about 6 spikes/stimulus for S+90 to about 3 spikes/stimulus for S-0, but does not go to zero as predicted. The noise response, although small, does show relative changes like that seen in the cross-correlator prediction, with the response for N+90 larger than that for N-90, which was larger than that for N0. The unit was therefore sensitive to azimuth for both signal and masker as predicted, but the relative rates for the signal and the noise were not as predicted. Consequently, the masking type for this unit was always suppressive, regardless of the position of the signal and masker. Additionally, the signal was detected in all of these conditions through an increase in rate, despite the fact that the cross-correlator predicted that the signal could not be detected for S0. In general, the difference between the model and neural response could be explained if something were increasing the signal response over the noise response.

Masking type index depends on noise azimuth

Throughout the sample of ITD sensitive units, the type of masking usually changes as a function of noise azimuth. To see how these changes relate to the best and worst azimuth, we plot the MTI as a function of both the noise azimuth (Figure 18, Panel A) and the noise relative IPD (Figure 18, Panel B). For these figures, we only show results for favorable signal azimuths ($|\phi_s| < 0.1$). When the noise is in the ipsilateral hemifield (negative azimuths), the masking is usually suppressive (MTI near -1). However, for the noise in the contralateral hemifield (positive azimuths), the noise can mask either through excitation (MTI near 1) or suppression (MTI near -1). This dependence on azimuth seems to arise from the fact that most of the units have their BITDs on the contralateral side. To examine how the positions of the best and worst azimuths affect the masking type, we also plot the MTI as a function of noise relative IPD. As discussed in the Methods section, azimuths near the best azimuth give relative IPDs near 0, while azimuths near the worst azimuth give relative IPDs near -0.5. Replotting these results in terms of relative IPD shows that the MTI changes abruptly around $\phi_n = -0.25$. When the noise is at an unfavorable azimuth, giving a relative IPD near -0.5, the masking is always

suppressive; however, when the relative IPD of the noise is favorable, giving a relative IPD near 0, the masking can either be excitatory or suppressive, despite the fact that the signal and masker are both at favorable azimuths. Because the individual unit responses are not always predicted by the cross-correlator model, additional processing beyond cross-correlation appears to affect how the noise masks the signal response.

Signal detection type depends on signal azimuth

As shown in the examples above, the signal can be detected through either an increase or a decrease in rate, and a cross-correlator can, but does not always, predict when this occurs. Figures 18, Panels C and D, show that the direction of the signal detection depends on both the signal azimuth and the signal relative IPD. For these panels, we only show results for favorable noise azimuths ($|\phi_n| < 0.1$). When the signal is at 0 degrees or in the contralateral hemifield (positive azimuths), the signal is detected through an increase in rate in most cases (108 out of 122 thresholds; Figure 18, Panel C). (Many of the points are plotted on top of each other, especially near 1.) The median SEI (red line) is near 1 in these cases. However, when the signal is placed at -90° , the signal is usually detected through a decrease in rate (11 out of 14). The SEIs never go to -1, but are usually near -0.5, indicating that the signal does not ever completely suppress the noise response at threshold. In Panel D, we plot the SEI as a function of the signal relative IPD. Placing the signal at a favorable azimuth ($\phi_s > -0.25$) almost always increases the overall rate (106 out of 122 thresholds), as expected, but occasionally decreases the overall rate. For signals at unfavorable azimuths ($\phi_s < -0.25$), the signal often decreases the overall rate as expected (9 out of 14 thresholds), but the signal is sometimes detected through an increase in the rate response. These results combined with the MTI results seem to suggest that some additional processing beyond that of a cross-correlator is affecting the relative rate of the signal and the masker.

Discussion

In this chapter, we showed how the responses and thresholds of low-frequency, ITD-sensitive single units in the inferior colliculus change when chirp trains and noise maskers are placed at different spatial locations, and we examined whether individual

unit responses showed correlates of spatial release from masking. We also compared the way the signal and masker responses change with azimuth to predictions from a cross-correlator model.

Because we are interested in the threshold for the most sensitive unit in the population, we found little advantage in using a method of signal detection that used the synchronized rate compared to the mean rate alone. The synchronized rate only improved the thresholds when the signal and masker were both excitatory, the condition that generally gave the worst thresholds. The synchronized-rate thresholds were actually worse than the rate thresholds in one of the most interesting cases, when the signal suppressed the noise response, although some other method of including the spike timing might improve these thresholds. Because the synchronized-rate and the mean-rate thresholds gave similar results in the case when the signal excites and the noise suppresses, which is the case where the best thresholds occur, the population thresholds shown in Chapter 4 are the same regardless of whether synchronized rate or mean rate was used.

We also showed that spatial separation of the signal and noise does not necessarily improve thresholds for single unit responses. Instead, the worst-threshold azimuths are correlated with the best noise azimuths, indicating that the worst thresholds seem to occur when there is a strong excitatory noise response. In Chapter 4, we show that the population of ITD-sensitive units shows a correlate of SRM even though individual units do not. Additionally, the best thresholds seem to occur when the signal is placed near the best azimuth, in contrast to the results shown by Jiang et al. (1997a, b). Our results do not appear to contradict these previous results because, when measured in one unit, we found results similar to those of Jiang et al. (1997a, b) when we used a pure-tone signal. By measuring the thresholds for the chirp train at different signal levels, we showed that the differences caused by the two signals were probably not due to the amount of energy through the peripheral filters, which would be different for these two signals, but is more likely to result from differences in their temporal or spectral characteristics.

As the cross-correlator model predicts, the noise response and the effect of the noise masker on the signal can change with noise azimuth. Similarly, whether the signal

is detected through an increase or a decrease in rate can change with signal azimuth. However, the neurons do not always behave as a cross-correlator model would predict: in some cases, the masker suppresses the signal response even for favorable noise azimuths, and sometimes the signal is detected through an increase in rate at unfavorable signal azimuths and a decrease in rate at favorable signal azimuths. For both the model and the data, the signal-plus-noise response was dominated by the signal-alone response at low noise levels and the noise-alone response at high noise levels. Incorrectly predicting the signal-alone rate compared to the noise-alone rate would explain these differences between the model predictions and the actual response.

That the relative rates of the signal and the noise are sometimes different from the cross-correlator predictions suggests that there may be additional neural processing beyond that of a cross-correlator. We usually used a chirp-train search stimulus for these experiments, which might introduce a bias into our sample for units that show an increased response to chirp trains. Because the chirp trains were designed to have a similar spectrum to the noise, the primary difference between the signal and the noise is in their timing characteristics. The signal is transient and has a strongly modulated 40-Hz envelope while the noise is continuous and not as strongly modulated. Therefore, one example of additional processing is that of neural adaptation, which could cause the response to the (transient) chirp train to be higher when the (continuous) noise level is low. Neural adaptation is evident even in the cross-correlator predictions of unit 22-0, which are based on an ANF model that includes adaptation; other units in the auditory pathway could show additional adaptation. However, adaptation is just one example of a neural mechanism that can give sensitivity to the stimulus envelope. Units in the IC have been shown to be sensitive to the stimulus envelope, with many units preferring 40-Hz modulation rates (e.g. Krishna and Semple, 2000). This sensitivity to the stimulus envelope could explain some of the differences between the cross-correlator predictions and the single unit data. In the next chapter, we implement a model that includes envelope processing to account for these differences and discuss possible neural mechanisms that could produce the sensitivity to the stimulus envelope.

In summary, we showed that for our sample of low-frequency, ITD-sensitive units, 1) single unit thresholds do not show correlates of SRM because separation

between the signal and masker does not necessarily improve the thresholds; 2) instead, the threshold appear to depend on the relationship of the signal and masker compared to the units' best and worst azimuths; 3) unlike previous physiological studies (e.g. Jiang et al, 1997) that used tones in noise in an MLD paradigm, the best thresholds for these stimuli occur for the signal at the best azimuth, and the difference between the studies seems to stem from differences in the signals; and 4) unlike the predictions of a cross-correlator model, the relative discharge rates of the signal and masker are not always predicted by their positions relative to the best and worst azimuths for the unit, indicating that additional neural processing changes the relative signal and masker responses. The subsequent chapters are devoted to developing a model of the individual unit's responses that includes both a cross-correlator model and an envelope processor (Chapter 3) and comparing the neural and model population responses to the psychophysical thresholds (Chapter 4).

Calretinin
Immunostain
(Lateral)

Nissl
Stain

Calretinin
Immunostain
(Medial)

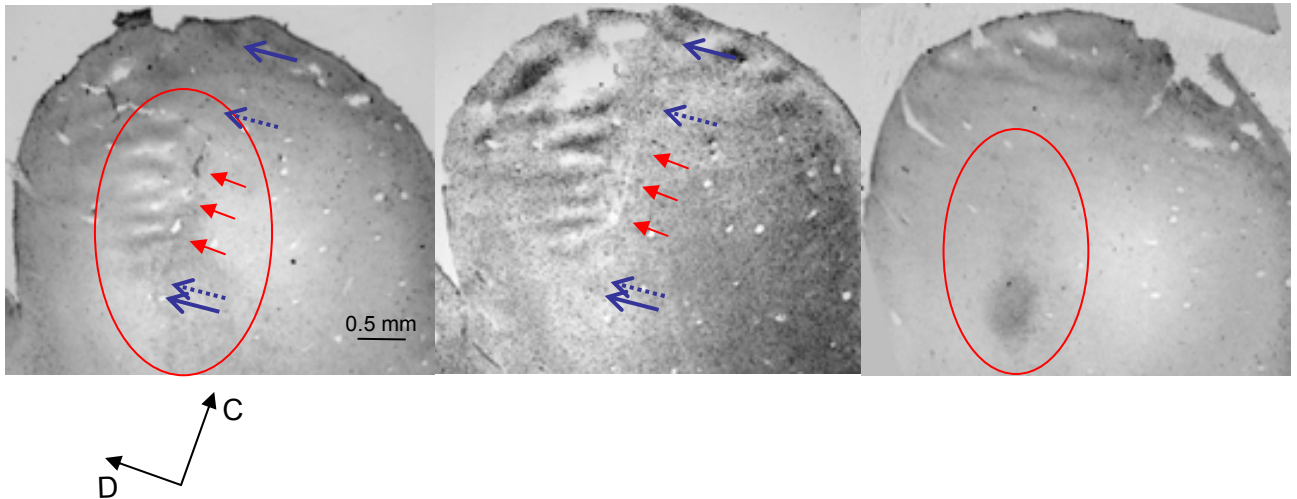


Figure 1: Histological sections for one animal. Left and right: Parasagittal sections immunostained for calretinin. Small, dark cloud of immunostaining apparent in right figure (more medial), larger more faint cloud marked by red oval in both panels. Red arrows in left panel point to electrode track. Solid blue arrows show approximately where stimulus-induced neural activity began and ended based on electrode depth measurements; dotted blue lines show where the unit recordings began and ended. Middle: Parasagittal Nissl-stained section showing electrode track (red arrows). Arrows same as other figures. Electrode depths are approximate as the starting point is not well defined, and any tissue shrinkage was neglected.

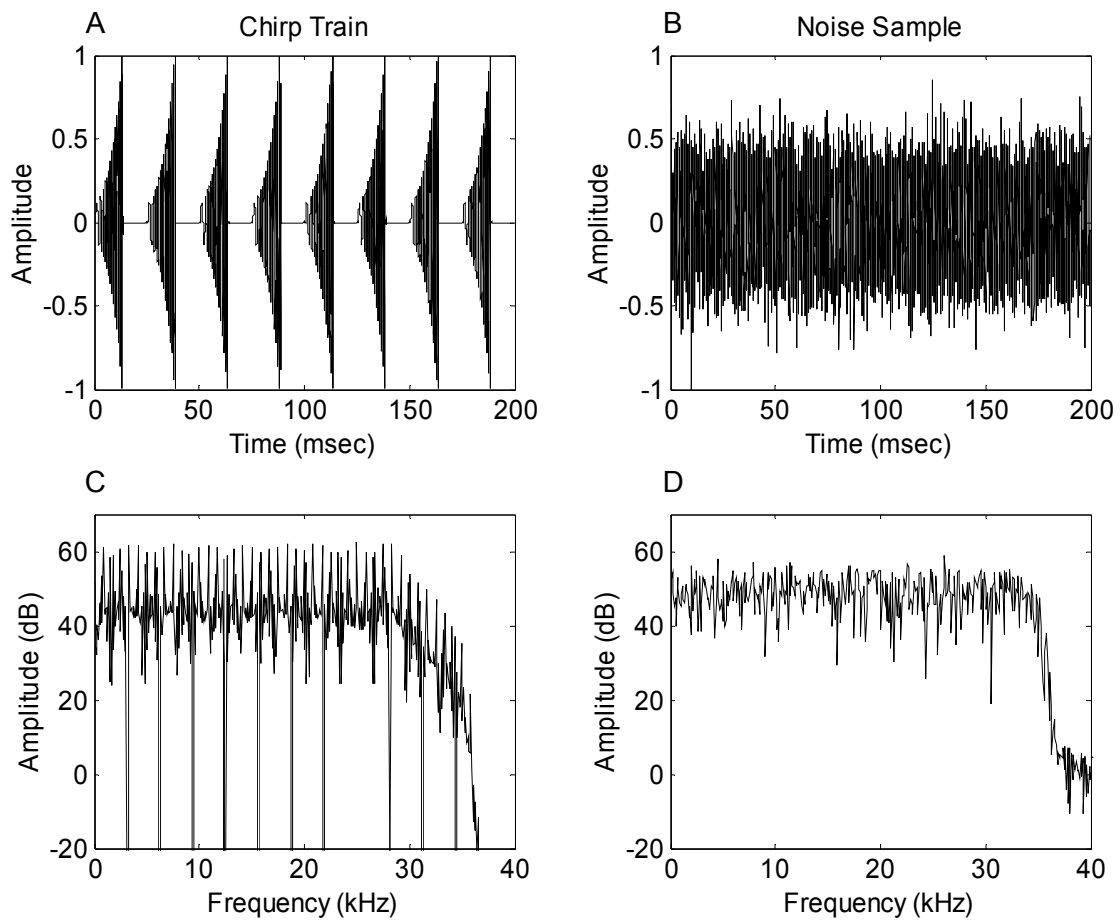


Figure 2: Broadband chirp-train signal (A and C) and broadband noise masker (B and D). A and B show the stimulus waveforms, and C and D show the corresponding spectra. The frequency of each chirp is swept from 300 to 30 kHz logarithmically; the envelope increases exponentially to flatten the spectrum.

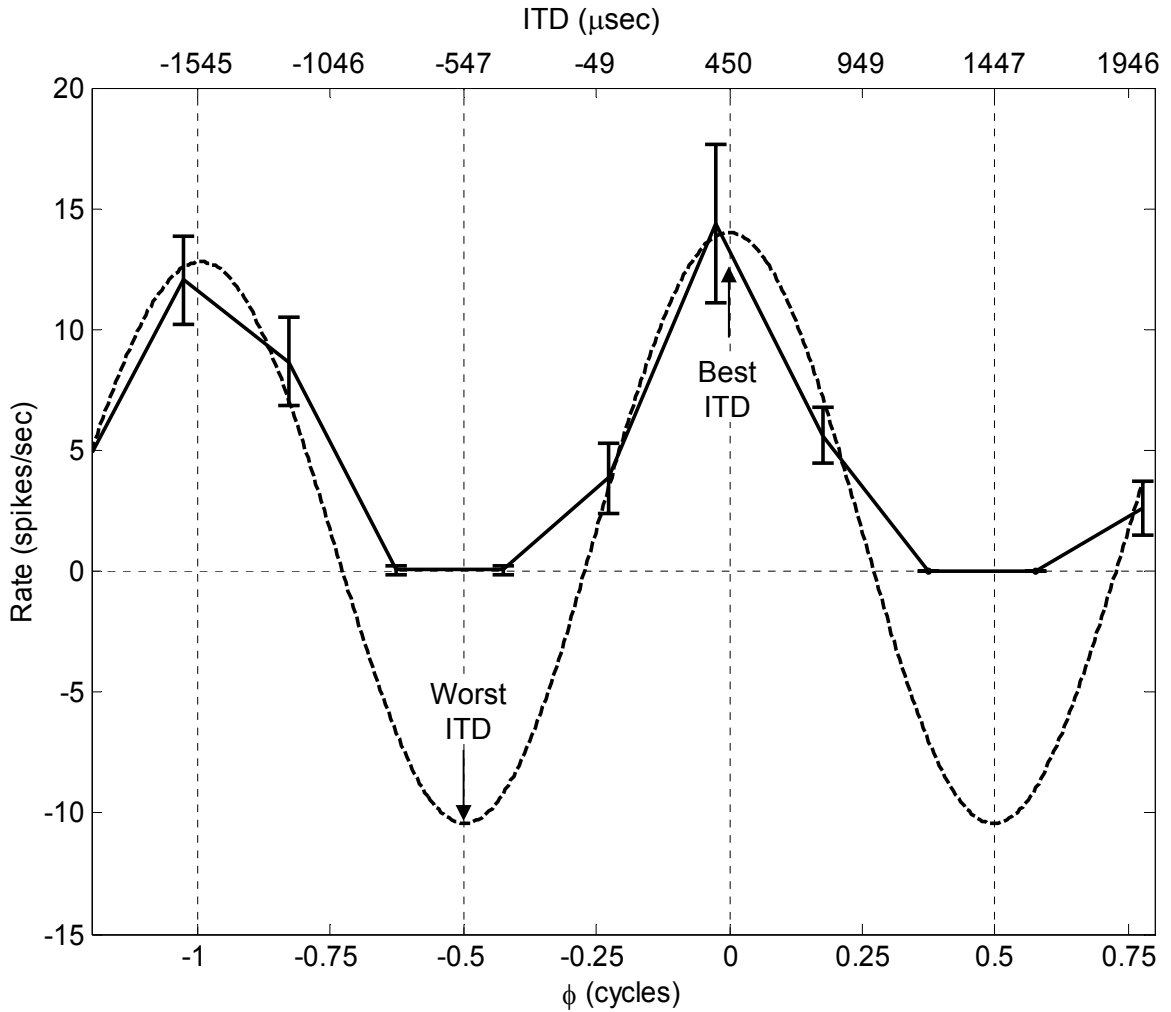


Figure 3: Gabor fit to noise delay function. Solid curve shows rate as a function of noise ITD (top axis) and relative IPD (bottom axis). The relative-IPD (ϕ) axis normalizes the response so that the best ITD (450 μ sec) is 0 cycles, and the worst ITD (-547 μ sec) is -0.5 cycles. The BF_{ITD} of this unit is 500 Hz. Dashed line is the Gabor fit without half-wave rectification to show the worst ITD.

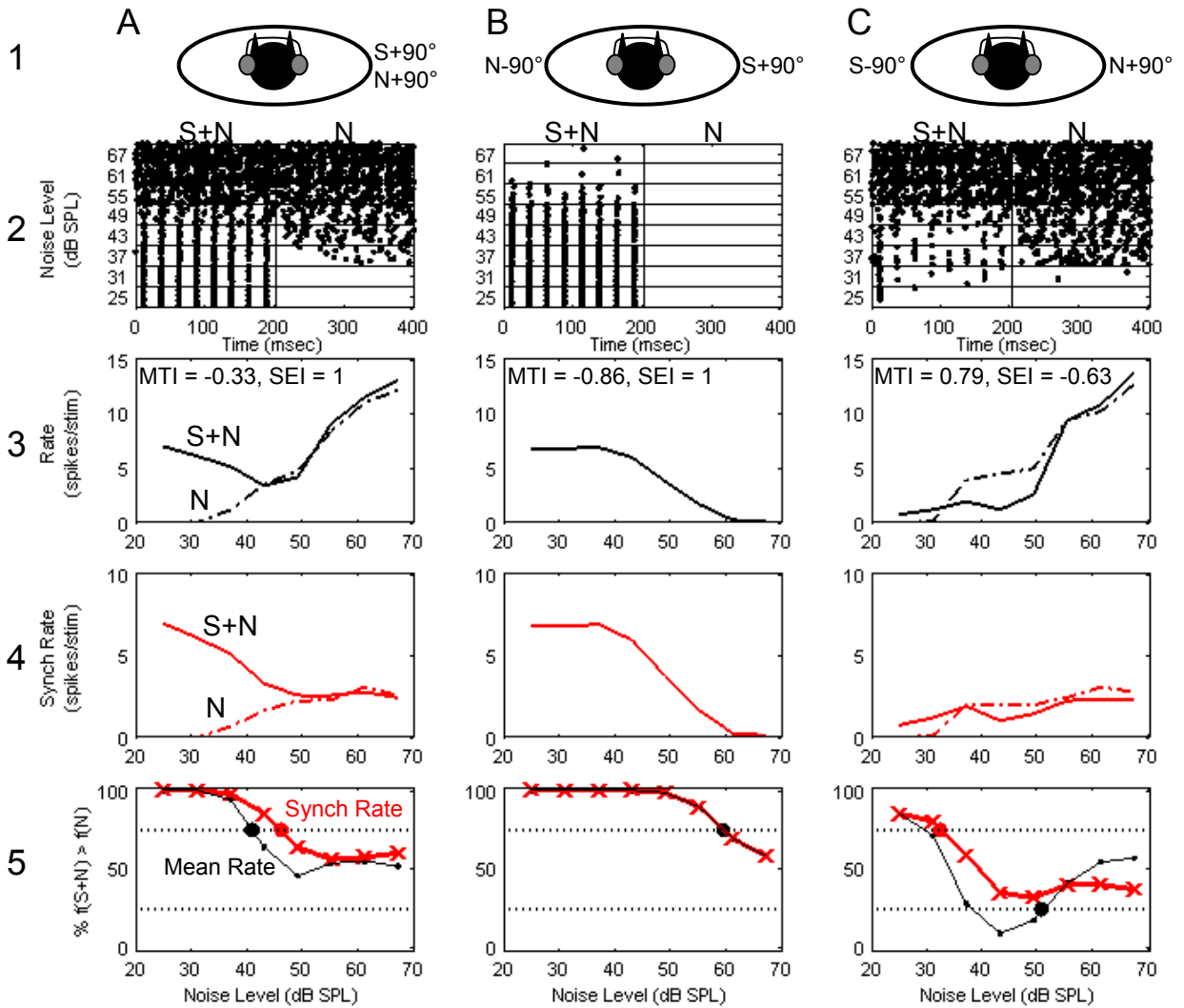


Figure 4: Response for a single unit (22-0: $BF_{ITD} = 743$ Hz, $BITD = 290$ μ sec) for three signal and noise configurations. The best azimuth for this unit is +90° ($\phi_{90^\circ} = 0.0$ cycles), and the worst azimuth is -90° ($\phi_{-90^\circ} = -0.43$ cycles). The signal level is 43 dB SPL. First row: configurations of signal and noise for the three columns, and, indicating that +90° is a favorable azimuth and that -90° is an unfavorable azimuth. Second row: Dot rasters as a function of noise level. Every dot represents a spike. Signal-plus-noise response in 0-200 msec, noise-alone response in 200-400 msec. The unit entrains to the chirp train at low levels (bottom of plots) and is masked as noise level is raised. Third row: Mean rate vs. noise level for signal-plus-noise (S+N, solid lines) and noise-alone (N, dash-dot lines). Fourth row: synchronization rate vs. noise level for signal-plus-noise (S+N, solid lines) and noise-alone (N, dash-dot lines). Fifth row: Percent of signal presentations where the rate (thin black lines with circles) or synchronization rate (thick red lines with x's) is larger for the signal-plus-noise window than the noise-alone window. Threshold is defined to be when the signal can be detected 75% of the time, corresponding to either 75% or 25% greater (dotted lines). Thresholds are noted by circles.

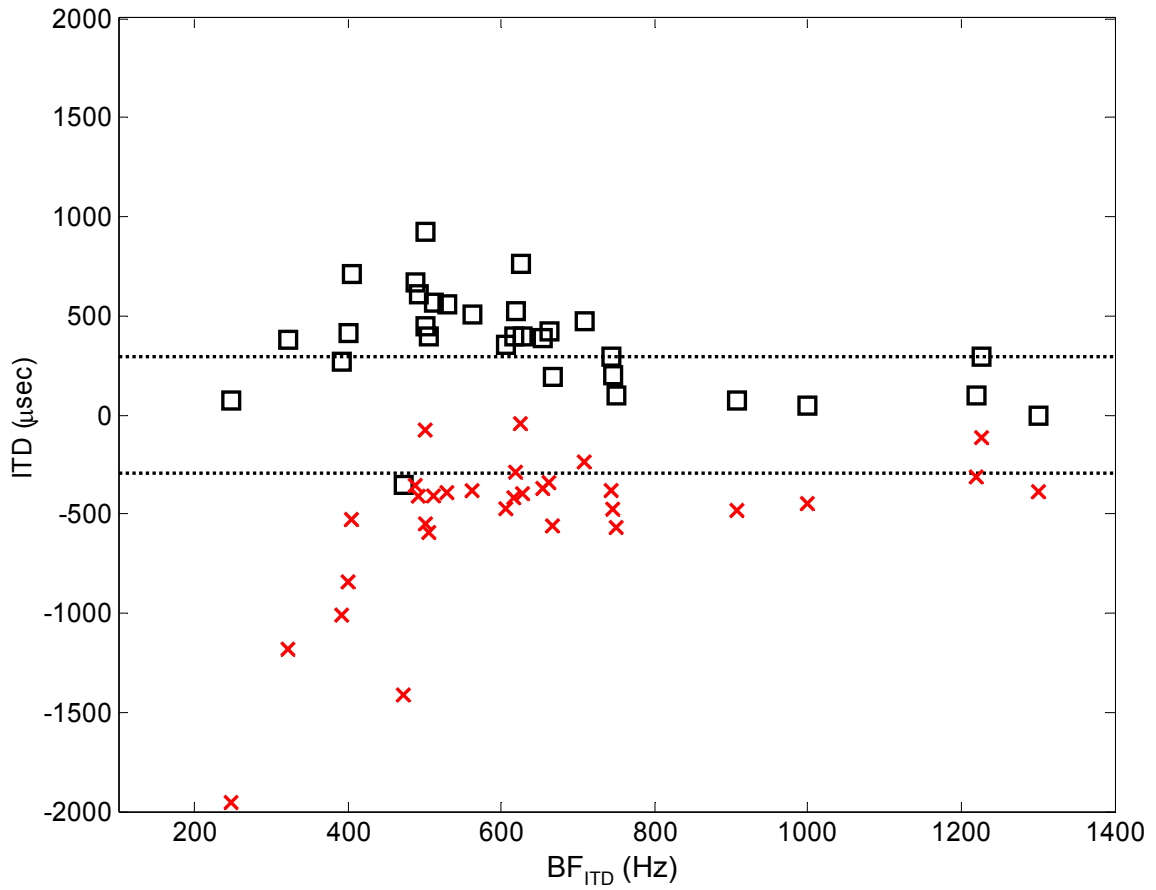


Figure 5: Best (squares) and worst (x) ITDs as a function of BF_{ITD} for all the units in our sample. Most of the best and worst ITDs lie outside the physiological range (dotted lines).

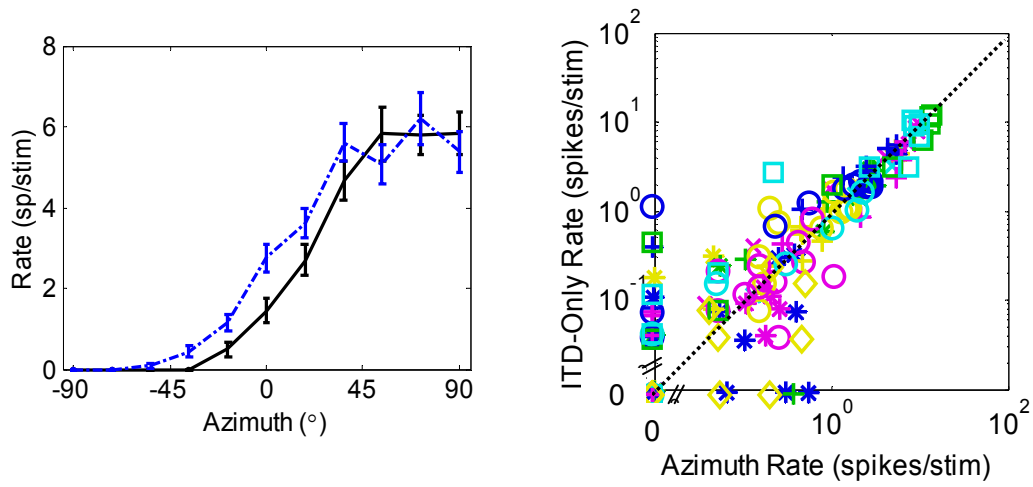


Figure 6: Azimuth sensitivity is due to ITD sensitivity. Left: Rate vs. noise azimuth (black, solid line) and rate vs. noise ITD (blue, dash-dot line) for same unit. Error bars show ± 1 standard error of the mean. ITDs were picked to match those in the HRTFs. Right: Noise-azimuth rate vs ITD-only rate for all the units in our sample for 19 units. Note log-log scale, showing that the rates are similar unless very small. Units on the axes have zero rate. Different combinations of marker color and shape indicate a different units.

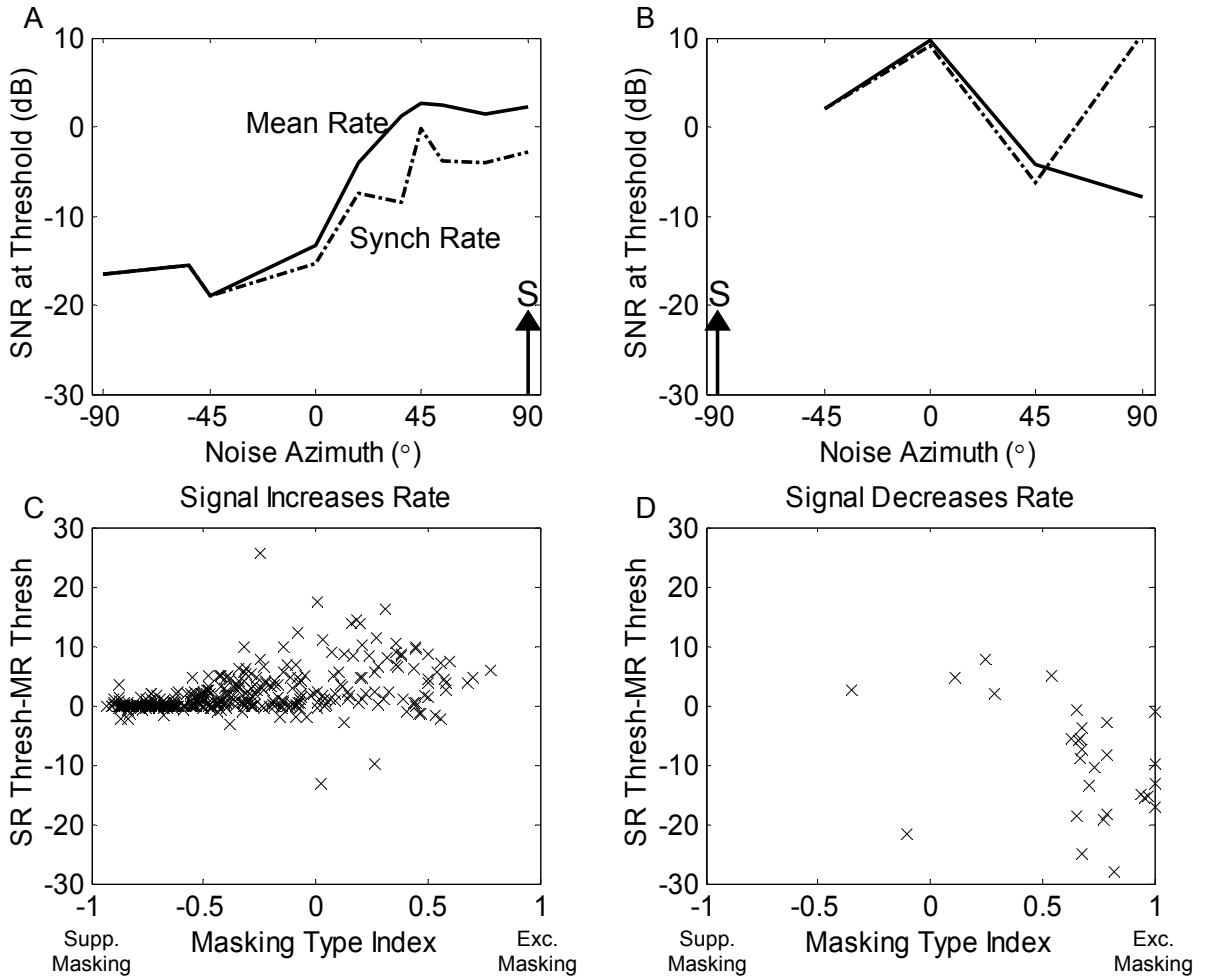


Figure 7: Mean-rate thresholds compared to synchronization-rate thresholds. A and B: Mean rate (solid lines) and synch. rate (dash-dot lines) thresholds for unit 22-0 (same as Figure 4) for S+90° (A) and S-90° (B). The synchronization-rate thresholds can be better or worse than the mean-rate thresholds, depending on how the signal is detected and the type of masking. C and D: When the signal is detected through a rate increase (C), the synchronization-rate thresholds are usually better or the same as the mean-rate thresholds, depending on the type of masking. When the signal is detected through a rate decrease (D), the mean-rate thresholds are usually better. The masking type cannot be close to -1 in this case as the unit would produce no response.

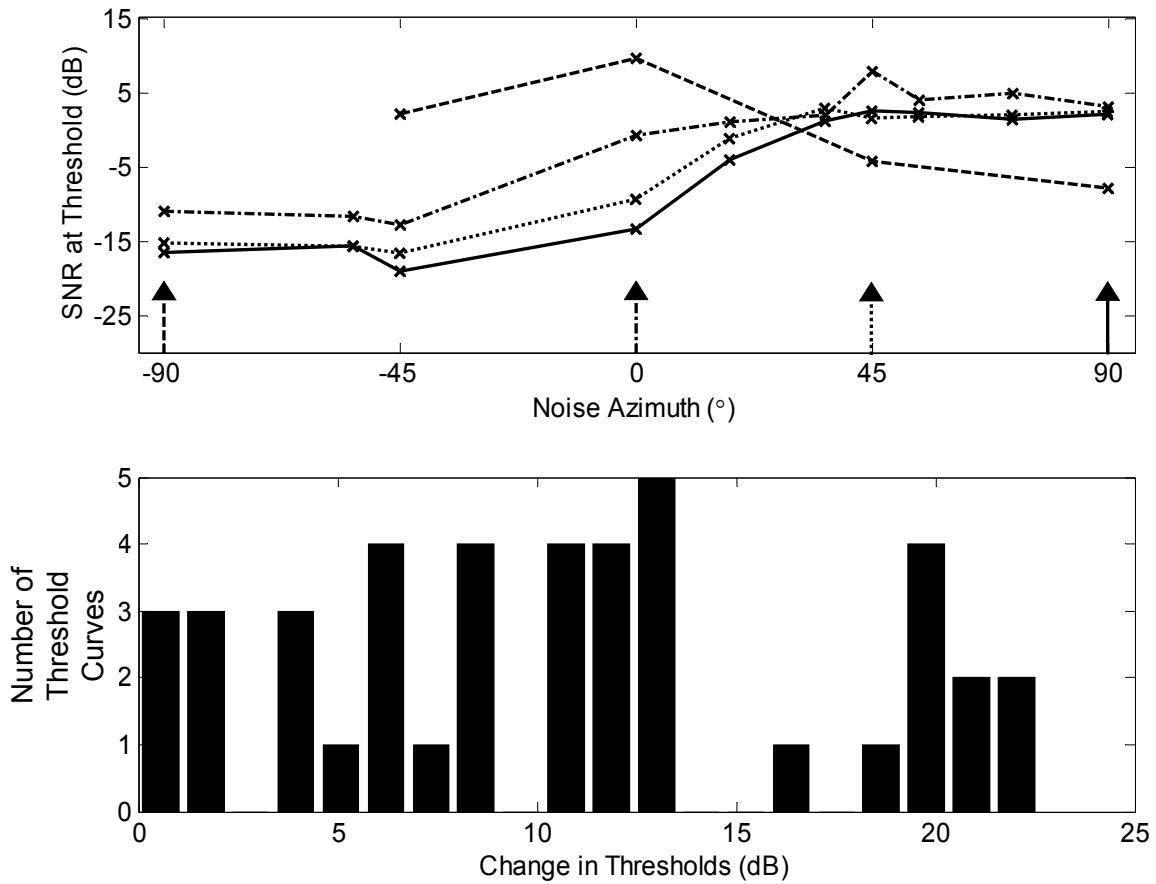


Figure 8: Top: Signal-to-noise ratios at threshold as a function of noise azimuths for four signal azimuths for unit 22-0 (Figure 4). The signal azimuths are denoted by the arrows, and the tails of the arrows indicate the corresponding threshold curve. Bottom: For each unit and signal response, the difference in threshold between the best threshold and the co-located condition for the signal at the same azimuth. In some cases, no spatial release from masking was seen; in others, the release could be greater than 20 dB. We included the threshold curves for all the signal azimuths measured for every unit so there are more points than units in our sample.

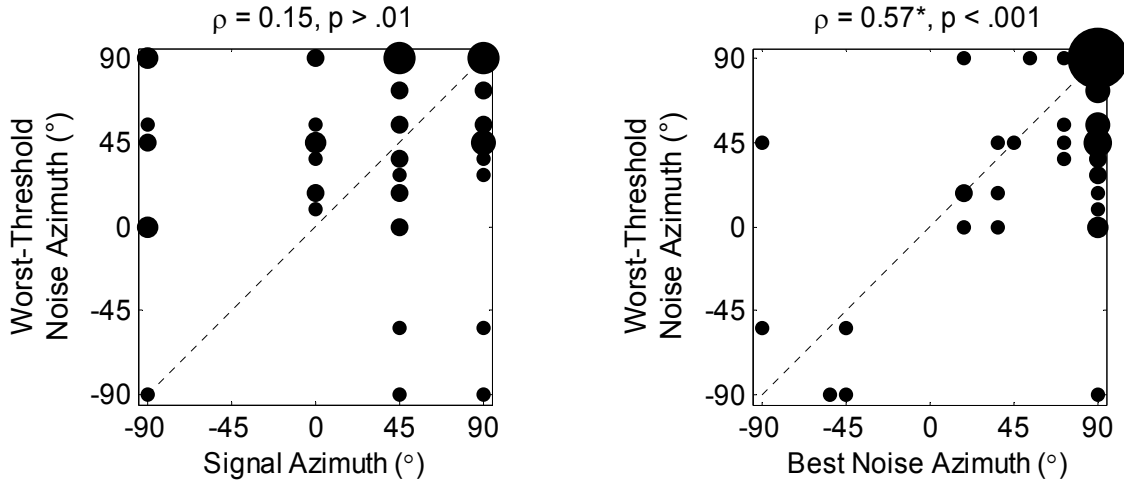


Figure 9: Bubble plot showing worst-threshold azimuth as a function of the signal azimuth (left) and the best azimuth (right, defined as the azimuth with the relative IPD nearest 0, right). The size of the bubble indicates the number of points at that graph location. The correlation between the most effective masking azimuth and the signal azimuth is 0.15, which is not significantly different from zero ($p > 0.01$), indicating that the worst thresholds do not necessarily occur when the signal and masker are co-located. The correlation between the most effective masking azimuth and the most favorable noise azimuth is 0.57, which is significantly different than zero ($p < 0.001$), suggesting that the worst thresholds occur when the masker response is largest.

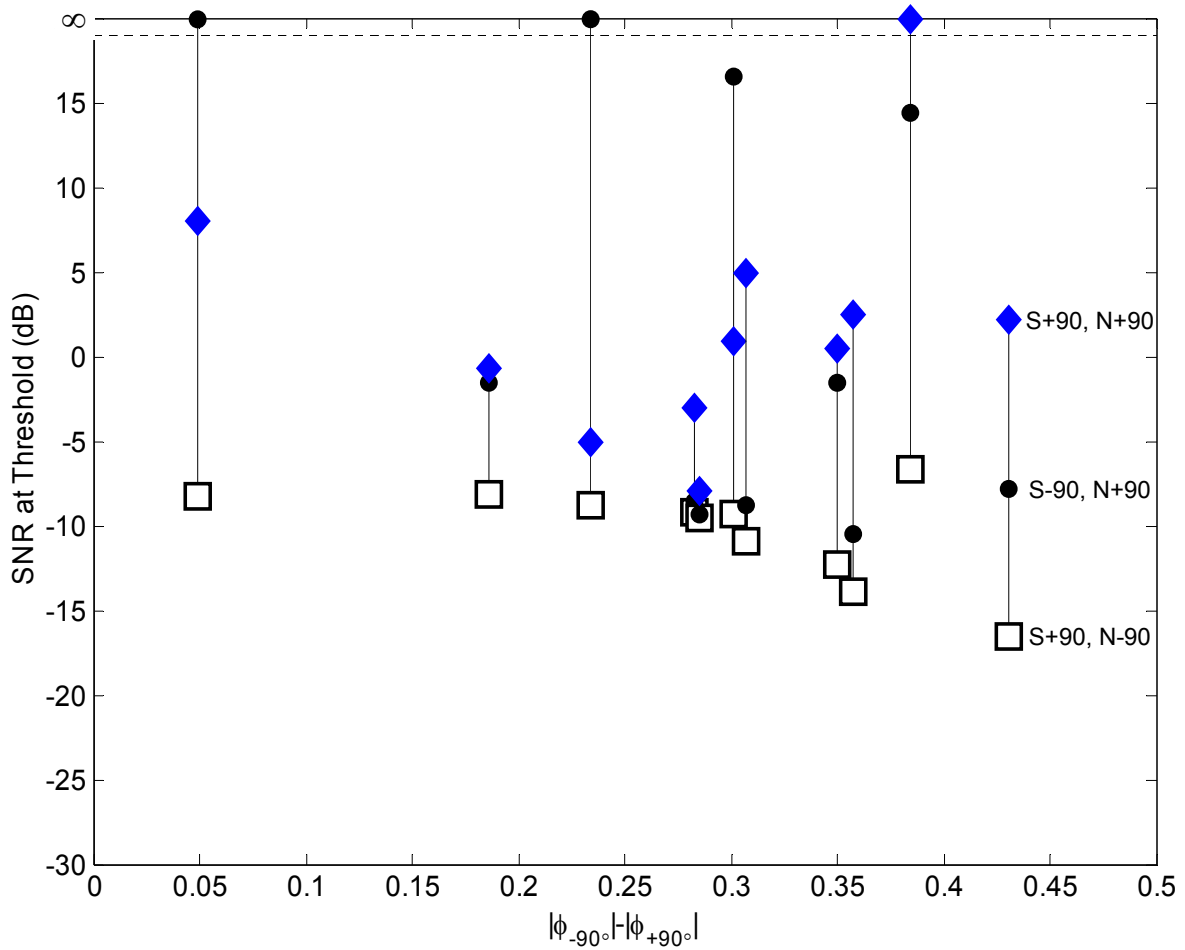


Figure 10: Masked thresholds for different units for S+90, N-90 (white squares), for S-90, N+90 (black circles), and for S+90, N+90 (blue diamonds). The x-axis is the difference in absolute value of the relative IPD at each unit's CF_{ITD} for -90° and $+90^\circ$. The S+90, N-90 thresholds are always better or the same as the other two conditions. The co-located condition (blue diamonds) was often, but not always, the worst condition of the three. Thresholds marked infinite could not be measured because the signal could not be detected at any masker level.

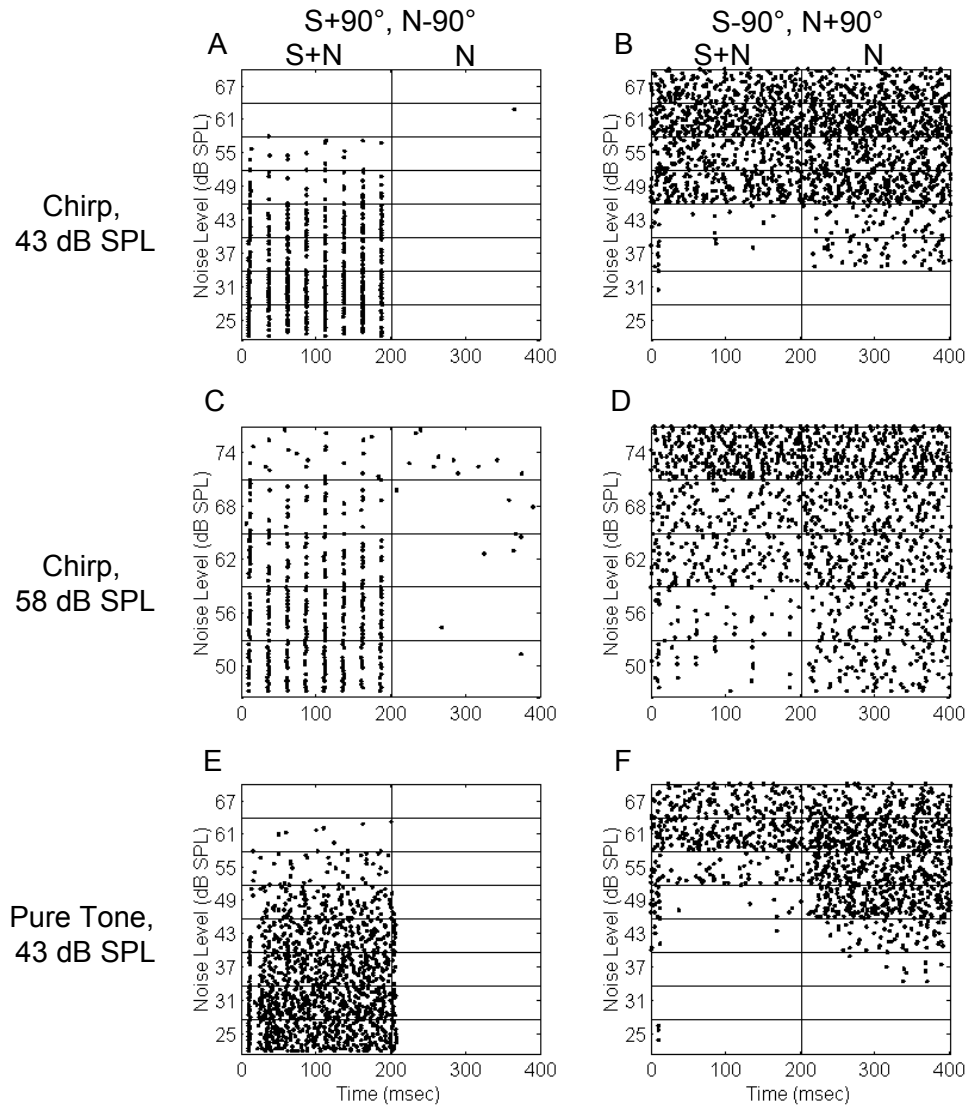


Figure 11: Dot rasters for the unit 26-5 (BF_{ITD} of 600 Hz, best ITD of 400 μ sec) for some of the conditions shown in Figure 12. A and B: Chirp train signal at 43 dB SPL for S+90, N-90 (A) and S-90, N+90 (B). The S+90, N-90 condition yields the better threshold. C and D: Same as top row, but for the signal at 58 dB SPL. As seen in Figure 12, the two conditions yield similar thresholds. E and F: Same as top row, but for a pure tone 500-Hz signal. Unlike the results for a chirp-train signal, the thresholds are better for the S-90, N+90 condition for a pure-tone signal.

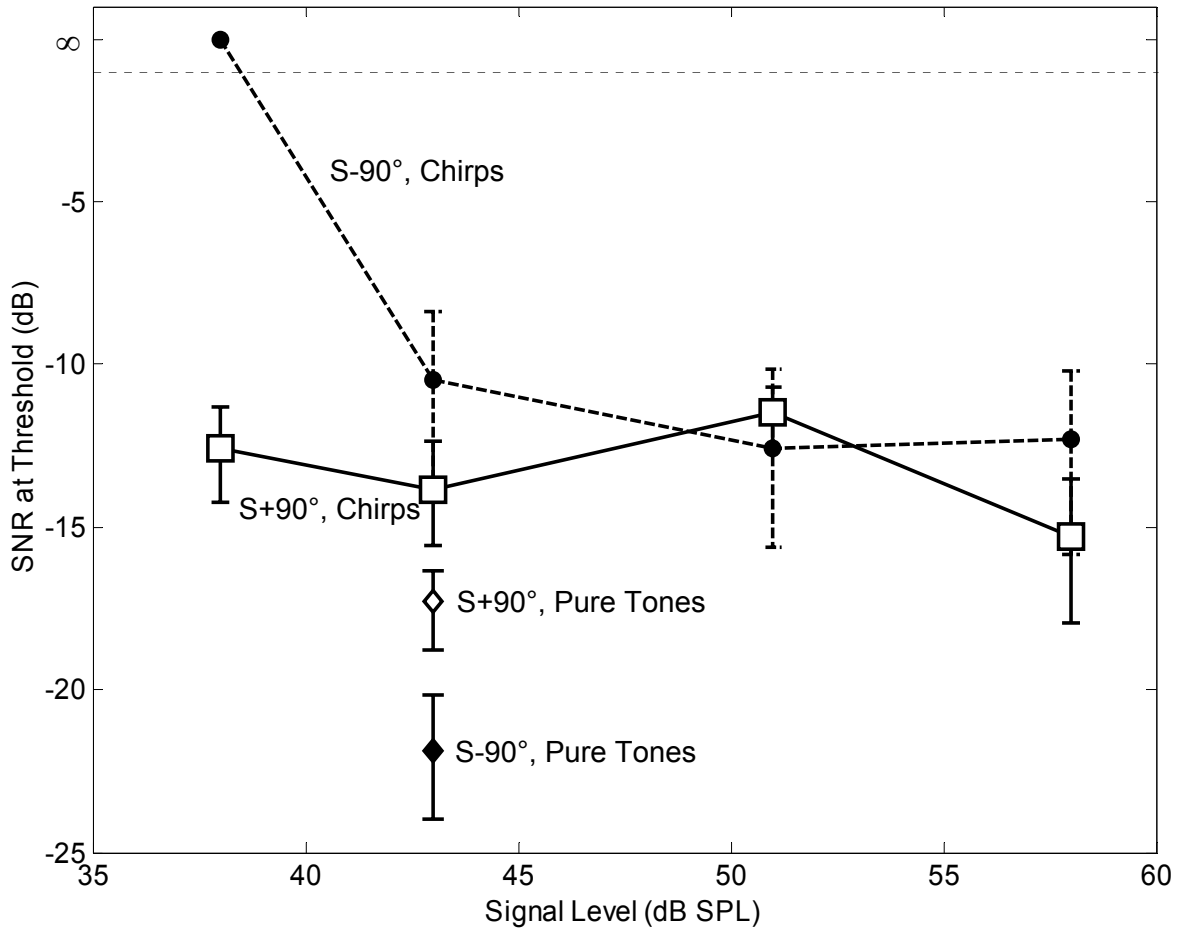


Figure 12: Masked thresholds for one unit (26-5: BF_{ITD} of 600 Hz, best ITD of 400 μ sec) as a function of signal level and type. Chirp trains at $+90^\circ$ (squares) remain fairly constant with signal level, but chirp trains at -90° (circles) cannot be detected at low signal levels. The -90° chirp thresholds improve with signal level to give thresholds similar to the $+90^\circ$ condition. A pure tone gives the reverse result: a 500-Hz tone placed at $+90^\circ$ (open diamond) gives a worse threshold than the same tone placed at -90° (filled diamond).

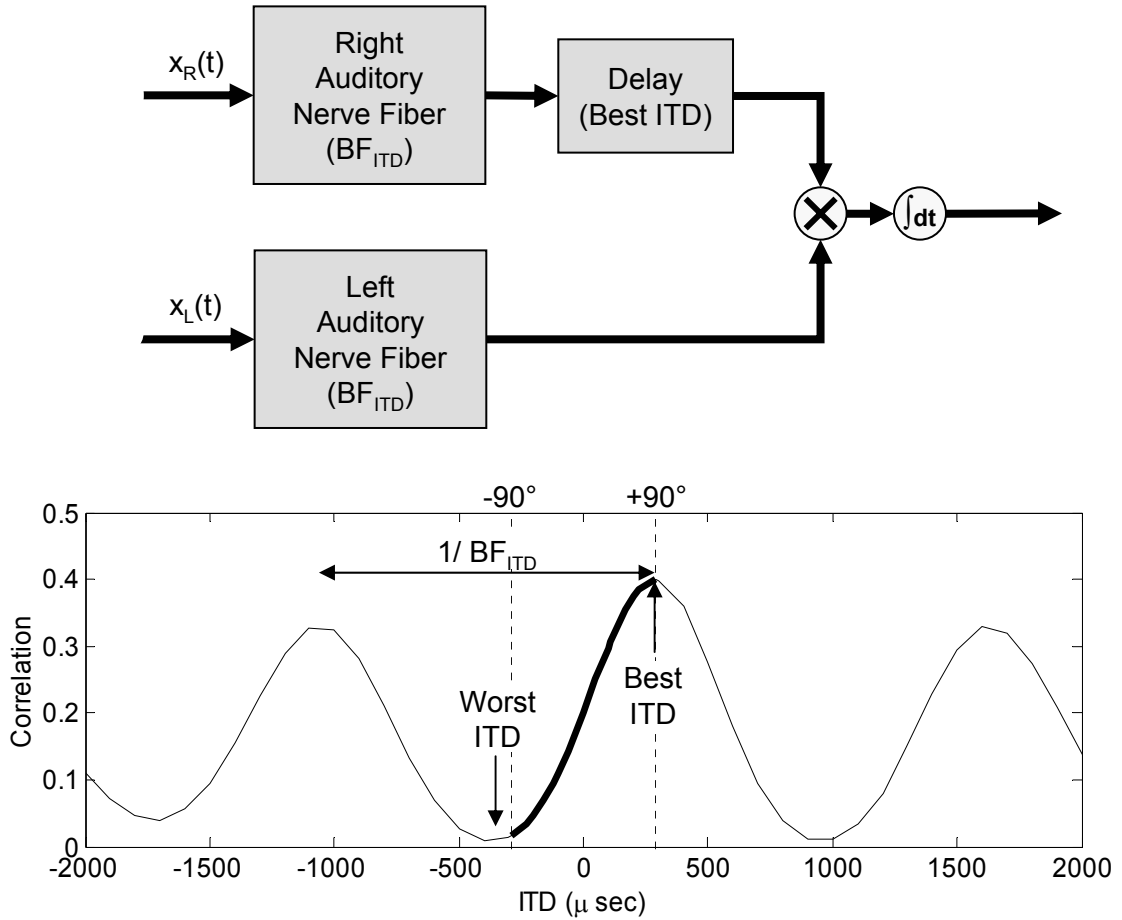


Figure 13: Top: Block-diagram of cross-correlator. The stimulus waveforms for the left and right ears, $x_L(t)$ and $x_R(t)$, are input to the auditory nerve fiber model (Zhang et. al, 2001), which outputs the probability of firing. Then output of the right auditory nerve fiber is delayed, resulting in a non-zero best ITD. The unnormalized correlation of the outputs is then computed. Bottom: The interaural correlation for a broadband noise for several ITDs (i.e., $x_L(t) = x_R(t-ITD)$). The narrowband filtering causes the correlation to be nearly periodic, and the peak location is determined by the best ITD. In this example the BF_{ITD} (743 Hz) and best ITD (290 μ sec) were those of the unit in Figure 4. The worst ITD is one-half period of the BF_{ITD} away from the best ITD. The physiological range, -290 to +290 μ sec as determined from the $+90^\circ$ and -90° degrees in our HRTFs, is marked by the vertical dotted lines.

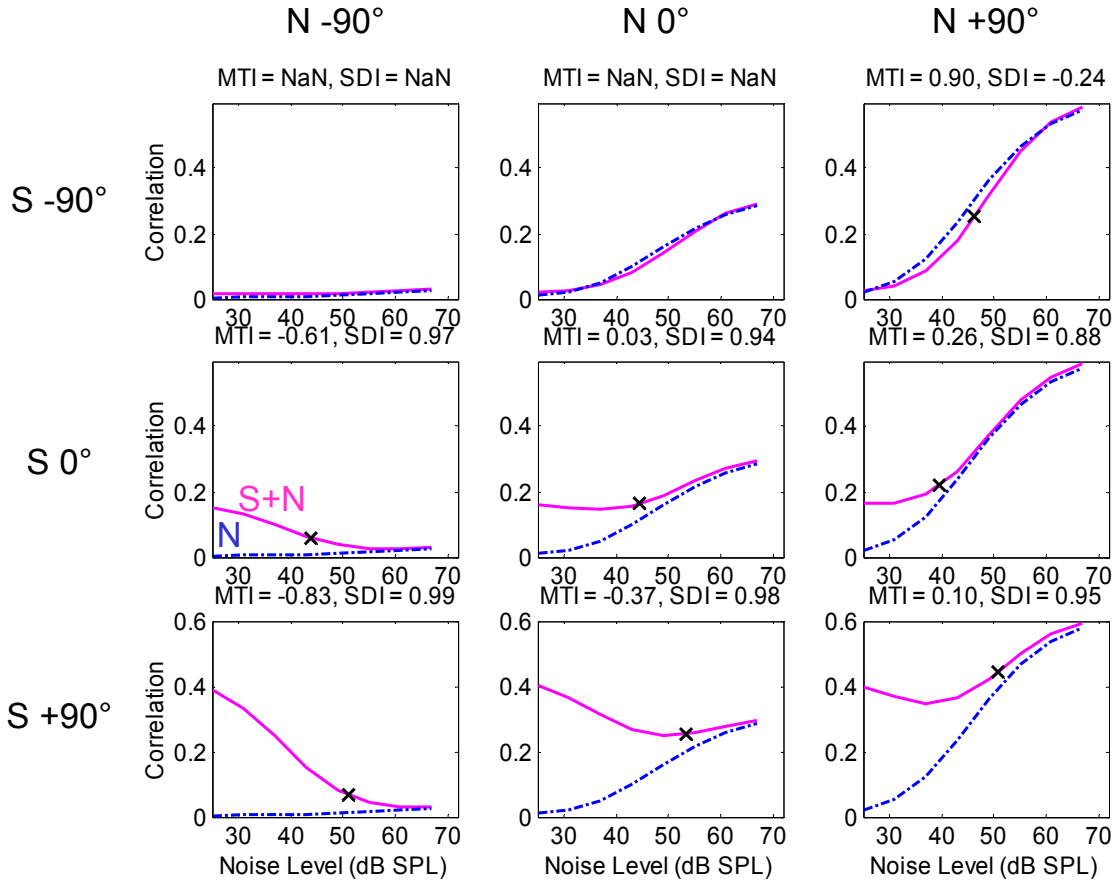


Figure 14: Cross-correlator predictions for unit 22-0 ($BF_{ITD} = 740$, best $ITD = 290 \mu\text{sec}$) with a best azimuth of +90 degrees and a worst azimuth of -90 degrees. Rows have a constant signal azimuth, and columns have a constant noise azimuth. Panels for co-located conditions are located on the diagonal beginning in the upper left corner. Magenta solid curve is the signal-plus-noise unnormalized correlation, and blue dash-dot curve is the noise-alone unnormalized correlation. The x's show the simulation threshold for a criterion of 0.05. The MTI and SEI for each condition are given in the titles.

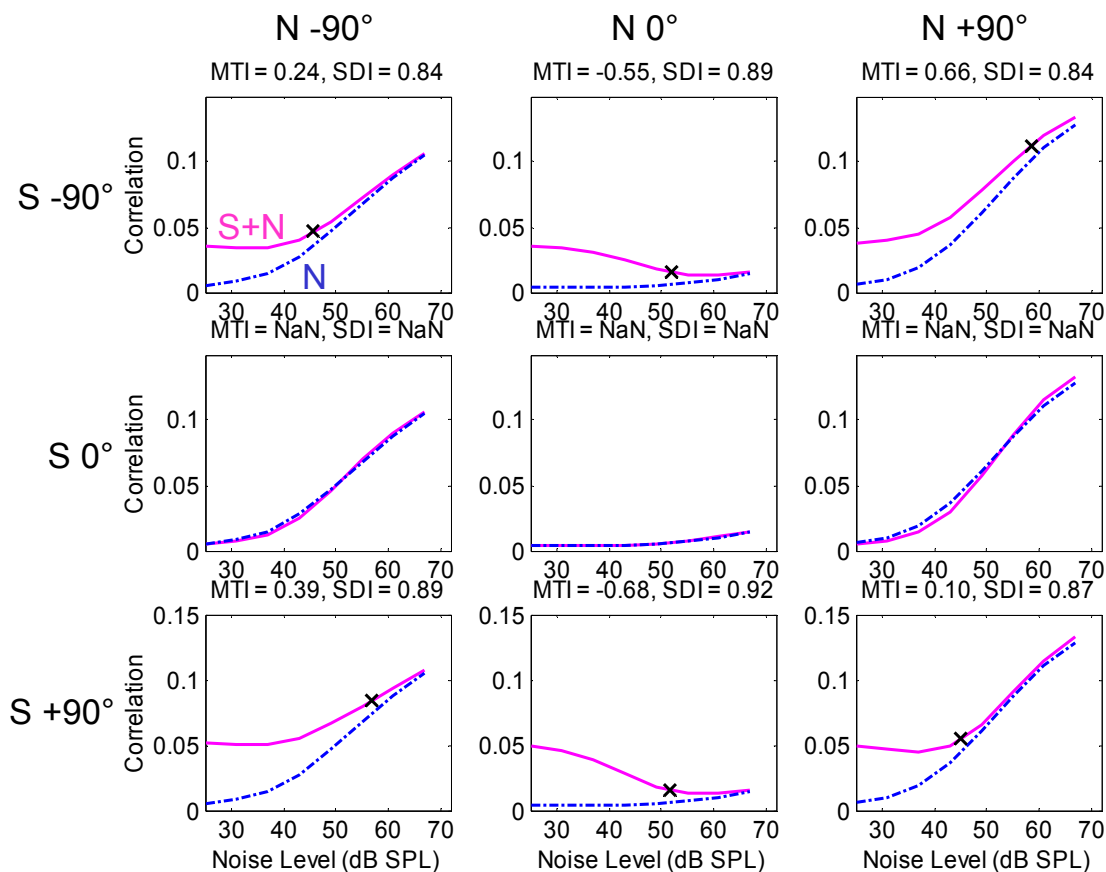


Figure 15: Cross-correlator predictions for unit 22-26 ($BF_{ITD} = 500$, best $ITD = 925 \mu\text{sec}$) with a best azimuth of $+90^\circ$ and a worst azimuth of -22° . Same format as Figure 2. The x's show the simulation threshold for a criterion of 0.01.

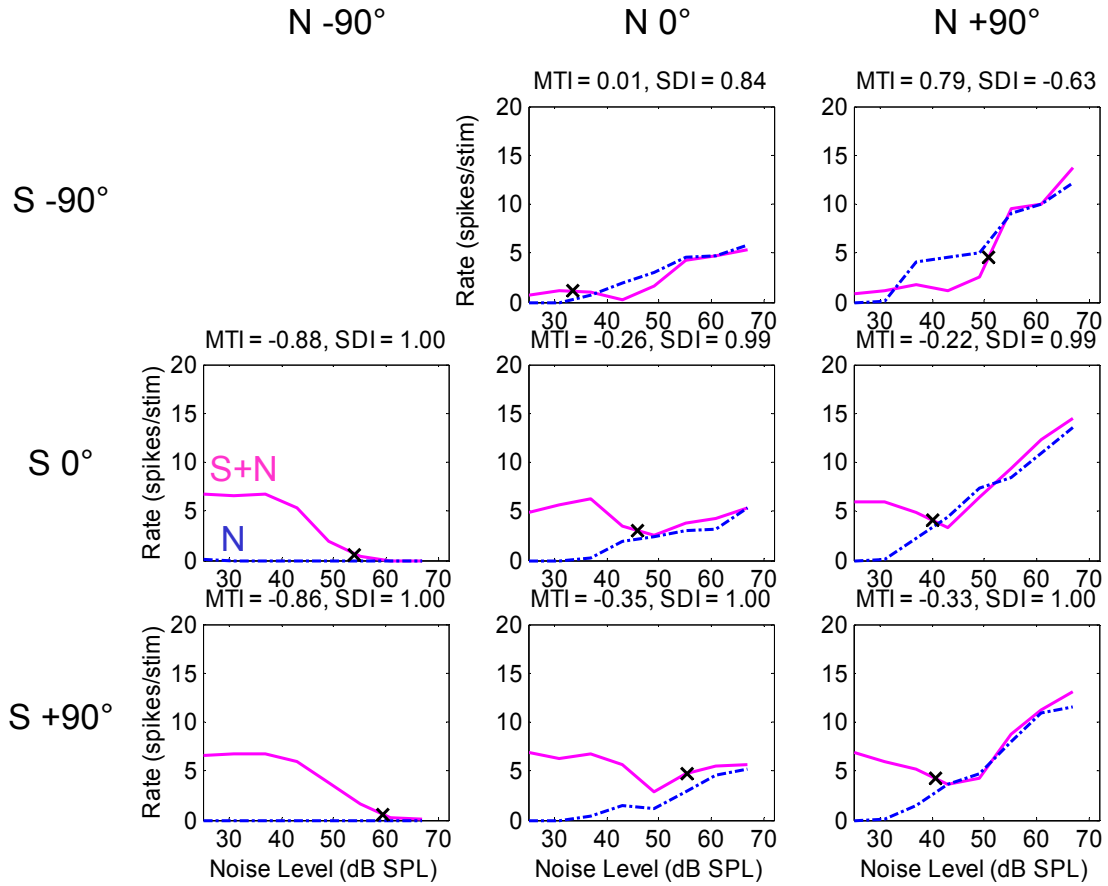


Figure 16: Actual rate responses for unit 22-0 ($BF_{ITD} = 740$, best ITD = $290 \mu\text{sec}$). Format similar to Figure 2. Magenta solid curve is the signal-plus-noise rate, and blue dash-dot curve is the noise-alone rate. The x's show the masked threshold, defined as described in the Methods section.

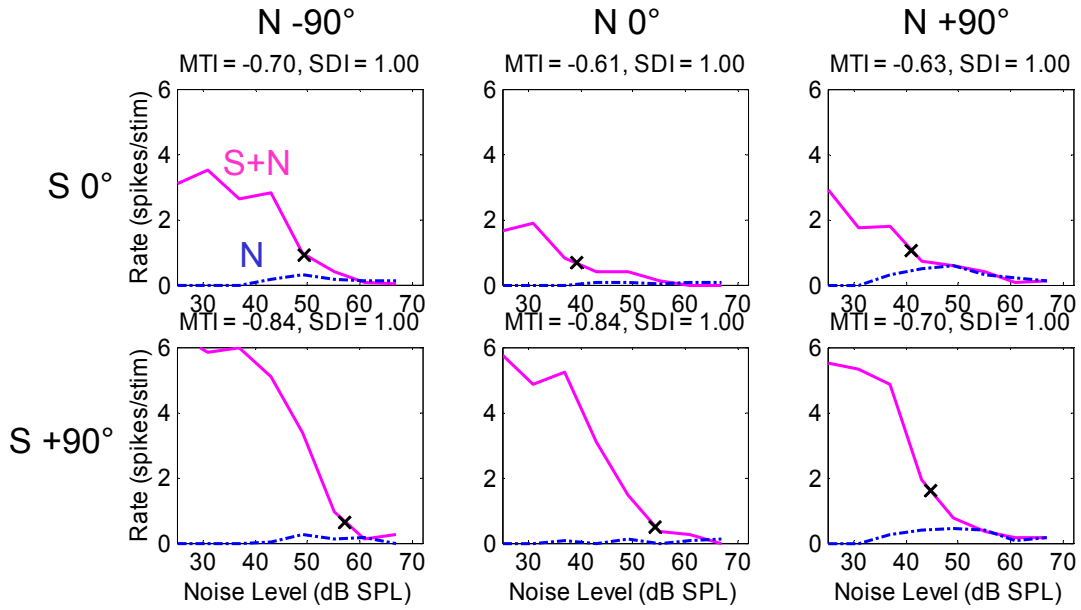


Figure 17: Actual rate responses for unit 22-26 ($BF_{ITD} = 500$, best ITD = 925 μ sec). Same format as Figure 3.

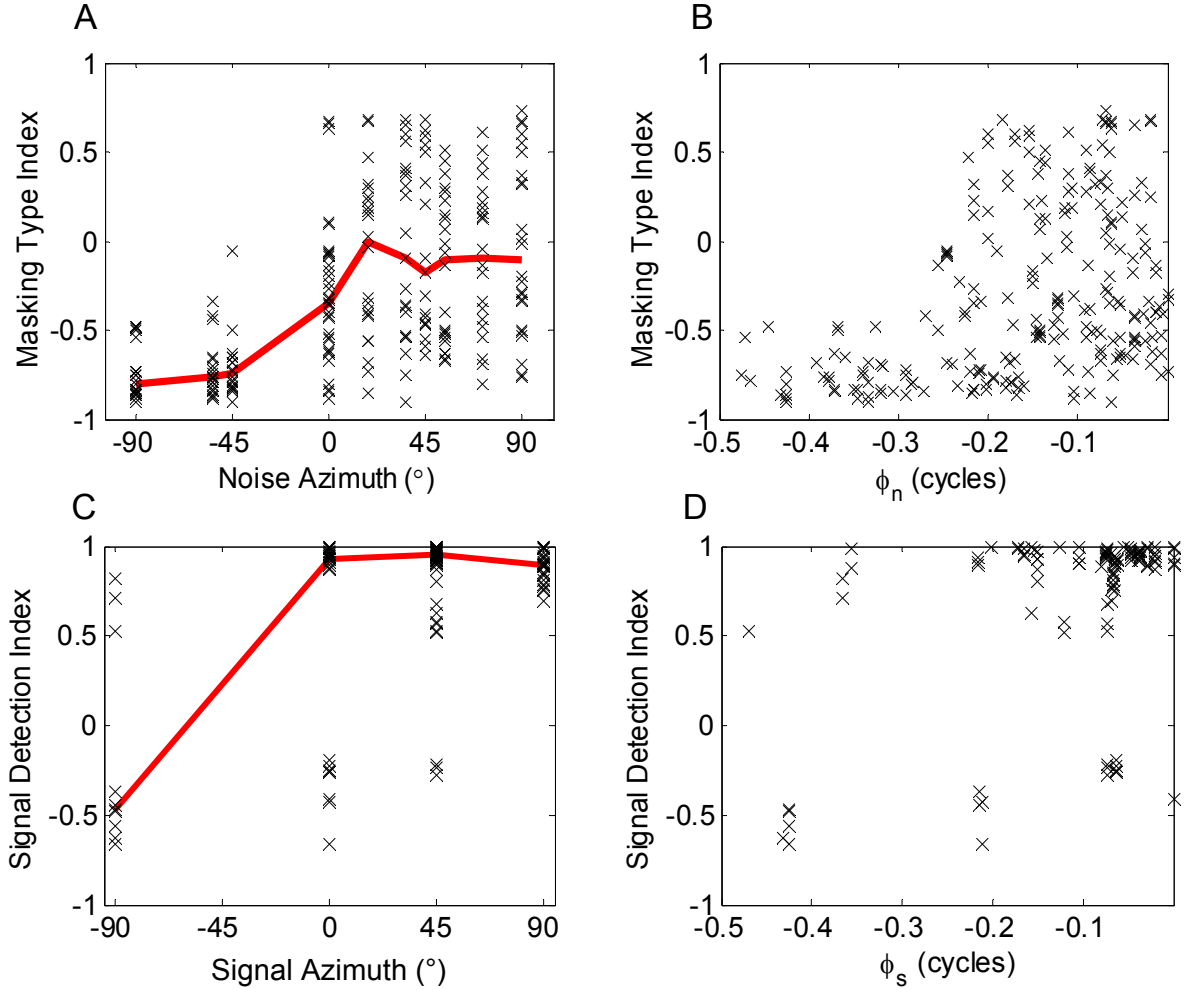


Figure 18: Masking type index (MTI) and signal effect index (SEI) change with azimuth and relative IPD. A) Masking type index as a function of noise azimuth for favorable signal azimuths ($|\phi_s| < 0.1$). Masking for negative noise azimuths (ipsilateral hemifield) yield mostly negative MTIs, indicating suppressive masking; positive noise azimuths (contralateral hemifield) give both positive and negative MTIs, indicating both excitatory and suppressive masking occur. Red line shows the median value. B) MTI as a function of noise relative IPD for favorable signal azimuths ($|\phi_s| < 0.1$). Unfavorable noise IPDs give suppressive masking, but favorable noise IPDs can give suppressive or excitatory masking. C) Signal effect index as a function of signal azimuth for favorable noise azimuths ($|\phi_n| < 0.1$). For positive signal azimuths, the signal is usually detected through increase in rate; for negative signal azimuths, the signal is usually detected through a decrease. Many of the points overlap because the signal was in the same location for multiple noise azimuths. Red line shows the median value. D) Same as C as a function of signal relative IPD for favorable noise azimuths ($|\phi_n| < 0.1$). The signal is usually detected through an increase for favorable IPDs and a decrease for unfavorable IPDs, but there are exceptions in both cases.

Chapter 3: A Computational Model of Single Unit Responses

Introduction

In Chapter 2, we showed how the responses of single ITD-sensitive units in the cat inferior colliculus (IC) change as signals and maskers are placed at different locations in space. Some of these unit responses were similar to those expected from a simple cross-correlator, but some differences were seen. Most notably, the relative responses to the signal and the noise alone were incorrect for some units because the signal response often seemed to be enhanced by additional processing not included in the traditional cross-correlator models. The additional processing could occur at many levels of the auditory pathway.

Because of this potential for additional processing beyond that of the cross-correlator, the purpose of this chapter is to clarify the processing required to predict the responses of the single units described in the previous chapter. Knowing the nature of this additional processing will give insight into the actual neuronal mechanisms involved in signal detection. Toward that end, we create a fairly simple model that predicts the rate responses and thresholds seen in our single unit data. We begin by building a model similar to the cross-correlator model described by Colburn (1973, 1977a, b). This model is assumed to reflect the simple coincidence detection that is thought to occur at the level of the MSO. We then add processing pathways that are tightly constrained by the data to account for the differences between the data and the model. This data-driven model shows what type of processing is required to explain the neuronal responses and thresholds. In determining the type of processing required, we will not necessarily restrict ourselves to physiologically viable mechanisms. However, the processing that is required is almost always consistent with known physiological mechanisms (see Discussion).

In this chapter, we first give an overview of the model. We then describe the development of the model, which evolves through comparisons of the model rate response to the single unit data. Finally, we compare the actual single unit thresholds to model thresholds and show the effects of varying the model parameters.

Model Overview

Figure 1 shows a block diagram of the model, which has three separate pathways to give ITD-sensitivity (white box), realistic rate-level functions (light gray box), and sensitivity to modulation frequency (dark gray box). The stimulus waveform from each ear is passed through the Zhang et al. (2001) auditory nerve fiber (ANF) model. The output of one of the ANFs is delayed so that the unit will respond maximally for a particular best ITD. The outputs of the ANFs then diverge, separating into the rate-level processor pathway and a cross-correlator pathway; the modulation-filter pathway branches off from the cross-correlator. The rate-processor and the cross-correlator are shown below to be effectively independent, and their outputs are multiplied. For the rate-level processor, the spike counts from the left and right auditory nerve fibers are averaged, and the average ANF rate is transformed via a memoryless input-output function to give the output rate.

The white box shows the path that provides the ITD-sensitivity by means of a cross-correlation of the ANF outputs. The output probabilities from the delayed ANFs are multiplied, integrated in time, and normalized by the geometric mean of the energies of the two input signals. These operations give the correlation coefficient of the two inputs at the delay corresponding to the model unit's best ITD. We then use another memoryless input-output function to transform the correlation coefficient to a scale-factor that ranges from 0 to 1. The outputs of the rate-level processor and the cross-correlator are then multiplied to give the unit's sensitivity to noise azimuth and level.

As shown in the Results section, the rate-level processor and the cross-correlator, which together constitute the “modified cross-correlator” model, are able to predict the noise-alone response accurately; however, the response to the signal is not well predicted by these two pathways alone. Because the signal and the masker have similar spectra, the poor prediction of the signal response is probably caused by the differences in the temporal properties between the signal and noise. For our experiments, the signal was a 40-Hz chirp train with a pronounced envelope modulation; the noise was continuous with no modulation envelope (although after peripheral processing, the noise will have some modulation). Consequently, we add a pathway that changes the overall rate based on the 40-Hz component of the multiplier output. Essentially, this pathway changes the overall rate based on the presence of the modulated signal. The measurement of the vector

strength (as defined by Goldberg and Brown, 1969) occurs after the cross-correlation because the signal-induced change in rate strongly depends on the signal azimuth. As we will show below, this model successfully predicts the single unit responses and thresholds for all of the units in our population.

Results

Rate responses

Noise-alone response

Before attempting to model the rate response of the neurons for the signal in noise, we model the noise-alone response. The noise-alone responses for two units are shown as a function of noise azimuth and noise level (Figures 2 and 3, Panel A). Consistent with the observations of McAlpine et al. (2001), the best and worst ITDs for both of these units are outside the physiological range (defined here by $+90^\circ$ and -90° , which corresponds to $+290$ and -290 μsec , respectively), giving monotonic rate-ITD functions within the physiological range. Since the ITD sensitivity largely determines the azimuth sensitivity for these low-frequency units (see Chapter 2), these units also have monotonic rate-azimuth functions. For the unit in Figure 2 (unit 22-0, CF of 740 Hz, best ITD of 290 μsec), the rate also increases monotonically with noise level, resulting in the maximum response occurring at $+90^\circ$ (contralateral to the recording site) and at the highest level, 62 dB SPL. For the unit in Figure 3 (unit 22-11, CF of 590 Hz, best ITD of 670 μsec), the rate-level function is non-monotonic, causing the maximum rate to occur at $+90^\circ$ and around 50 dB SPL.

A cross-correlator model similar to the one described by Jeffress (1948) and Colburn (1973, 1977a, b) is an obvious starting point for modeling these low-frequency ITD-sensitive units. Such a model can produce ITD sensitivity similar to that seen in single unit responses in the IC (see Irvine, 1992, for a review) and MSO (Goldberg and Brown, 1969; Yin and Chan, 1990). The cross-correlator pathway in Figure 1 shows our implementation of the cross-correlator. The stimulus waveforms to the two ears $x_L(t)$ and $x_R(t)$ are filtered through the Zhang et al. (2001) auditory-nerve fiber (ANF) model for a fiber with a spontaneous rate of 50 spikes/sec. This model includes narrowband filtering

due to the cochlear mechanics, rectification like that seen in the receptor potentials of the inner hair cells, a low-pass filter due to the membrane capacitance of the inner hair cells, and the dynamics of synaptic vesicle release; this model also incorporates appropriate processing for wide-band stimuli. The output of the ANF model is the instantaneous firing rate as a function of time. To obtain the model responses, we use the same stimuli as those used in the physiology experiments, including the same noise samples. Because the ANF model includes adaptation (due to the dynamics of synaptic vesicle release), we simulate the effects of our continuous noise stimuli by first playing the noise alone, followed by the signal-plus-noise, then followed again by the noise alone response. Only the response for the second noise-alone window is used. The results presented are for the average of 16 stimulus presentations. The output of the right auditory nerve fiber is then delayed to give the model unit a best ITD, and the correlation ρ of the output of the two auditory nerve fibers is computed. So that we can match the single unit responses shown in the previous chapter, the best ITD and CF for the model are set to the values determined by the Gabor fit to the noise-delay function (see Chapter 2). At this point, we are presented with our first real modeling choice: the correlation can either be normalized by the geometric mean of the energy in the input, giving the correlation coefficient that ranges from -1 to 1 (0 to 1 in this case, as our inputs are never negative), or left unnormalized. As we stated in the Introduction, we will use the data to decide how to proceed. Panels C in Figures 2 and 3 show the unnormalized correlation as a function of noise azimuth and level, and Panels F show the correlation coefficient. Both give appropriate sensitivity to noise azimuth, but without normalizing the correlation, this processing pathway depends on the overall level of the stimulus; with normalization, the dependence on overall level is largely removed. For the unit in Figure 2, the unnormalized correlation appears to predict the noise-alone response fairly well (Panel C compared to Panel A). However, the unnormalized correlation does not predict the response to unit 22-11 as well (Figure 3, Panels A and C). In this case, the unnormalized correlation cannot predict both the monotonic rate-azimuth functions and the non-monotonic rate-level functions shown in Figure 3, Panel A: a change in azimuth that increases the correlation causes an *increase* in rate, but a similar increase in correlation due to a change in level would cause a *decrease* in rate. Overall, neither the normalized

nor the unnormalized correlation alone gives satisfactory results because neither is able to completely explain both of the single unit responses shown.

At least two variables, then, must be employed to account for the dependence on noise azimuth, θ_n , and noise level, L_n . Because the simplest case would be if the unit responded to changes in these variables independently, we test whether the responses to noise azimuth and level are independent (in other words, if the response matrix is separable). If so, the overall rate is the product of a term dependent on level and a term dependent on azimuth:

$$R(L_n, \theta_n) = R(L_n) A(\theta_n)$$

If the responses are separable, the rate-level function and the rate-azimuth function can be computed independently and multiplied to produce the output rate. Because we have the complete rate response as a function of noise level and noise azimuth, we can use singular value decomposition (SVD) to test for separability (Pena and Konishi, 2001). Such analysis gives the functions for $R(L_n)$ and $A(\theta_n)$ that minimize the squared difference between the predicted response and the actual response. For these two cases, the rate response appears to be separable because the predicted response closely resembles the actual response (Figures 2 and 3, Panel B compared to Panel A). More generally, Figure 4 shows the noise-alone response compared to the response predicted by the SVD analysis for all the units in our sample (left). The center line indicates the identity line, and the outer lines show the average standard deviation for the neural rate response, which increases with overall rate. Overall, the differences between the predictions and the actual responses are usually less than the variations in the data; however, for low data rates, the prediction error can be larger than the variability in the data. The amount of variance accounted for by the SVD predictions as a function of the maximum noise-alone rate is shown on the right panel of Figure 4. The SVD synthesis accounts for a very large fraction of the variance seen in the data. As seen in the previous panel, at low rates, separability does not seem to hold as well, perhaps because we have a poorer estimate of the rate when the neurons fire less frequently. The fact that the rate response is usually separable suggests a model structure for which the noise level and noise azimuth are processed independently, and the results of the two processing paths are multiplied, as shown by the white and medium gray pathways in Figure 1.

From this analysis, we now know what the output of each processing path should be and how the output of these processing paths should be combined. We have not yet described, however, these separate pathways. The desired output functions for these two processing paths should match the functions $R(L_n)$ and $A(\theta_n)$ as noise level and azimuth are varied, but neurons do not have direct access to the noise level and azimuth. They only have access to the processed outputs of the auditory nerve fibers (ANFs). Therefore, we chose two ways of processing the auditory nerve fiber outputs that could give reliable and independent estimates of the noise azimuth and level: the normalized correlation ρ as discussed above and the average firing rate from the two ears r . These two variables are shown for the two example units in Figures 2 and 3, Panels F and H respectively. As shown above, the normalized correlation is monotonically related to the noise azimuth and nearly independent of the noise level; conversely, the ANF average rate is monotonically related to the noise level and independent of the noise azimuth. The average rate and normalized correlation, however, are not our only options: any other reliable and independent estimates of the noise level and azimuth could also serve as inputs. We can now replace $R(L_n)$ and $A(\theta_n)$ with $R(r)$ and $A(\rho)$. These input-output functions are shown in panels D and E in Figures 2 and 3. The correlation input-output function, $A(\rho)$, was fit using a half-wave rectified parabola and a threshold, and the rate input-output function, $R(r)$, was fit by two connected lines with different slopes and a threshold by minimizing the squared error using Matlab's *lsqnonlin* function. Note that the unit in Figure 3 (Panels A and E) has a non-monotonic input-output function for $R(r)$ that produces the necessary non-monotonic rate-level function. The noise-alone responses for the modified cross-correlator model for both units are shown in Figures 2 and 3, Panel G. Although the fits to the input-output functions smooth out some of the variation seen, the model does predict the data fairly accurately. Figure 5 plots the actual vs. predicted rates for all of the units in the population. For rates above a few spikes per stimulus, the model responses are usually within the range of variation in the data. At lower rates, the model tends to do less well, in part because the neuron responses are not separable in this case. Figure 6, Panels A and B show the rate and correlation input-output functions used to model all the units in our population (Panel C will be described below). The correlation functions have been multiplied by the maximum noise rate for

that unit so that the effects of the functions can be compared across units. The thresholds and shapes of the curves vary widely, and many of the rate input-output functions are non-monotonic. As shown above, the traditional cross-correlator model using only the unnormalized correlation would not have predicted the responses for any of the units with non-monotonic rate input-output functions.

Signal-plus-noise response for the modified cross-correlator

In this section, we examine the effect of adding the signal to the noise for both the modified cross-correlator model and the data. We first show the responses and the predictions for one example unit for the signal placed at -90° , which is fairly well predicted by the model. We then show the responses and modified cross-correlator predictions for two example units' responses for the signal at $+90^\circ$. The signal-alone responses are generally not well predicted by the modified cross-correlator model, as expected from the model results shown in Chapter 2. To better predict the data, we modify the model so that the response depends on the modulation envelope of the stimulus. We present the predictions for this new envelope-processor model and then discuss the implementation details.

Figure 7 shows the response of unit 22-0 (the unit in Figure 2) for the signal at -90° . When placed at this unfavorable azimuth, the signal suppresses the response to the noise (black arrow, compare to the noise-alone response in Figure 2, Panel A). The modified cross-correlator model does a fair job of predicting this response, in that the model also predicts a suppression of the noise response; the suppression for the data is somewhat larger than the predicted suppression for the model.

Figures 8 and 9, Panel A show the signal-in-noise response for the signal placed at $+90^\circ$ for units 22-0 and 22-11, the same units shown in Figures 2 and 3. As mentioned above, $+90^\circ$ is the best azimuth for both units. The response to the signal alone is apparent in the increased rate at low noise levels (thick edge and arrow) compared to the noise-alone response in Figures 2 and 3, Panel A. As the noise level is raised, the rate response approaches the noise-alone response, masking the response to the signal. As evident in unit 22-0's response (Figure 8, Panel A), the noise can mask either through

suppressing the signal response (for the noise at -90°) or by overwhelming the signal response (for the noise at $+90^\circ$).

The modified cross-correlator model predictions for this condition are shown in Figures 8 and 9, Panel B. For the model predictions, the signal-alone responses, indicated by the thick edge and arrow, are not as large as they are for the actual data for these two examples. Figure 10 compares the signal-alone rate (that is, when the noise is at its lowest level) to the noise-alone rate for all units with the signal and noise at the same azimuth and level. For the data (Panel A), the signal rate is rarely the same as the noise rate and is usually larger. For the modified cross-correlator model (Panel B), the signal rate is slightly lower than the noise rate in all cases. This difference between model and data is actually worse than that seen for the model in Chapter 2 because for the modified cross-correlator model, we use average rates and normalized correlations. Consequently, the effects of adaptation which are included in the ANF model are removed, making the signal and the noise alone responses nearly the same in all cases.

Signal-plus-noise response for the envelope-processor model

Because the modified cross-correlator model does not successfully predict the signal-in-noise response, we look at the error in the model to find a solution. Figures 8 and 9, Panel D show the difference between the actual response (Panel A) and the modified cross-correlator's response (Panel B). The difference is generally largest at low noise levels and decreases as the noise level increases, indicating that the differences occur when the signal response is relatively strong. As suggested by Figure 10, it seems that there is a difference in the way the signal and the noise are processed that is not included in the modified cross-correlator model. To create a change in the overall rate when the signal is present, we need a measure that can separate the signal from the noise and change the response when the signal is present.

In order to change the response for the signal without changing the response to the noise, we add a processing path that changes the overall rate if the output of the multiplier has a strong 40-Hz component, which is the modulation rate of the signal (Figure 1, dark gray box). This new model, which includes all the blocks shown in Figure 1, is termed the “envelope-processing model”. For this model, the cross-correlator

and the rate-level processor work as previously described. However, the 40-Hz vector strength is computed on the output of the multiplier. This vector strength is then transformed using another memoryless input-output function to give a signal-induced change in rate (“boost”) that is added to the modified cross-correlator response. (Although we call the signal-induced change in rate a “boost”, for some units, the change is negative.)

The envelope-processing model’s predictions for our example units can be seen in Figures 7, 8, and 9, Panel C. The boost improves the predictions of the modified cross-correlator model, particularly the signal-alone rate compared to the noise-alone rate for the signal at $+90^\circ$ (Figures 8 and 9). In these cases, unlike the cross-correlator model, the signal-alone rate (indicated by the thick edge and arrow) are comparable to the data. Figure 10, Panel C shows that the envelope-processing model allows the signal and noise to give different rates, unlike the modified cross-correlation model, and gives relative rates for the signal and masker similar to those seen in the data. Figure 12 shows the actual rate compared to the predicted rate for all the units and signal-masker combinations for the cross-correlator model (left) and the envelope-processor model (right). The envelope-processor model predicts the responses better than the modified cross-correlator model, although the errors are larger for the signal-in-noise condition than for the noise-alone condition in Figure 5. Furthermore, the noise-alone responses are largely unchanged because the boost is usually near zero for small vector strengths (Figure 6, Panel C). As we saw for the SVD analysis and the noise-alone response, error in the model predictions is generally within the variability of the data for high rates, but at low rates, the error in the predictions within the variability of the data.

Envelope-processor model implementation

The implementation of the envelope-processor model is fairly straight-forward, but there are a few important details. First, we measured the 40-Hz vector strength after the multiplier in the cross-correlation because the sign and magnitude of the boost varies with signal azimuth, as seen by the fact that the predictions at -90° were better than those for $+90^\circ$. As shown in Figure 11, the 40-Hz component for the model auditory nerve fiber response (Panels A and B) is similar for both $+90^\circ$ and -90° because the effective

level change due to varying the signal position is only a few dB at these low frequencies. Determining the appropriate boost is much easier based on the MSO responses (Panels C and D) because the vector strength of these responses changes with azimuth.

To obtain the input-output function for the boost, we compare the vector strength to the desired boost. We derive a sign for the vector strength from the phase: if the signal is suppressing the noise response (Figure 11, Panel C), the overall response has a π phase shift compared to when the signal excites the neuron. Phases like those obtained for the signal exciting the neuron give positive vector strengths, and the others, when the signal is suppressing the response, give negative vector strengths. The cut-off phase for the vector strengths that are made negative is a parameter of the model; see below for details on how this parameter was fit.

The difference between the cross-correlator model and the data is plotted in Figures 8 and 9, Panel D, and the differences between the data and model is plotted as a function of the vector strength in Figures 8 and 9, Panel F. There are many sources for the scatter in the plot, including the noise in the data, the possibility of not correctly estimating the parameters of the modified cross-correlator model, and the vector strength not being the correct input to use. Given all these sources of error, it is remarkable that a trend can be seen.

To fit a curve to the rate change, we used an equation of the form

$$\Delta R = A(1 - e^{(-vs - vs_0)/\tau})$$

where vs is the 40-Hz vector strength, A is the amplitude, vs_0 is the threshold vector strength, and τ is the time constant. This function is half-wave rectified and is fit separately to the positive and negative vector strengths. The results of all of these fits for our population can be seen in Figure 6, Panel C. The shapes of these curves vary widely from unit to unit. For large vector strengths, when the signal response is strong, the boost is usually positive, but it can also be negative. This finding suggests that while the signal alone usually causes an increased rate response over the noise-alone response, the signal can also cause a decrease in the rate. As the vector strength goes to zero, that is, when the signal response is small, the desired boost also goes to zero because, by construction, the model for the noise-alone response does not have a consistent bias. Finally, when the change in vector strength is negative, indicating that when the signal suppresses the noise

response, the change in response can be zero, positive, or negative. The corresponding changes in rate are not as large for negative vector strengths compared to positive ones, indicating that for the signal at an unfavorable azimuth, the desired boost is smaller than when the signal is at a favorable azimuth.

There are some obvious outliers in the lower right corner (in the oval) of Figure 9, Panel F. For these points, even though the vector strength was large, indicating a strong signal presence, there was little boost required for the modified cross-correlator model to match the data. These points occur when the noise is at unfavorable azimuths and suppresses the signal response. A possible explanation for these points could be that the mechanism involved in changing the rate for the signal response may not be activated if the inputs to the mechanism are sub-threshold. There are two different thresholds of this type in the model: the level threshold and the correlation threshold. Figure 3, Panel D shows no response for input correlations below about 0.3, and Figure 3, Panel E shows no response for ANF rates below about 20 spikes per stimulus. Because the signal level and azimuth are fixed, the ANF rate is guaranteed to remain above the average rate threshold, but when the noise is at an unfavorable azimuth, the overall correlation could be pulled below the correlation threshold ρ_0 . If we force the boost to be zero in such cases, we can predict the “outliers” in Figure 9, Panel F.

This small model detail, although affecting only a few azimuth and level combinations, can affect the unit thresholds, as described below. There are several parameters that we have mentioned above: the cut-off phase for deciding which vector strengths should be negative, the parameters of the positive and negative exponential fit, and the correlation threshold for the boost. In order to fit these parameters, we estimated initial values for these independently and gave these initial values to Matlab’s *lsqnonlin* function to provide the best fit (the one that minimizes the least squared error) for both the signal-plus-noise and noise-alone responses. Overall, this model successfully predicts the responses of many ITD-sensitive, low-frequency units with only three processing paths: a rate-level processor, a cross-correlator, and a modulation filter. Although there are several parameters involved for each processing path, we will show in the Parameters section that, in general, the thresholds are only sensitive to a few of these. First, however, we show that this model can also predict the actual thresholds.

Model thresholds

Since we now have a model that predicts the neural rate responses, we would like to compare the thresholds from this model to the individual unit thresholds. In Chapter 2, we defined individual unit thresholds in terms of the percent of stimulus presentations for which the signal could be detected. Our model responses, however, are deterministic: they yield the same response given the same input. Although there is some external noise due to the different noise samples, the model lacks internal noise. Without internal noise, the model thresholds are often much better than those seen in the data. We therefore added internal noise to give more realistic predictions of the thresholds. This internal noise model is a zero-mean Gaussian noise sample added to the overall rate. The variance of the noise was determined based on a linear fit to the variance of each unit individually as a function of the noise-alone rate. Because the response to the signal was highly synchronized to the modulation envelope of the signal, the signal-alone response usually had a much lower variance than the corresponding noise response; therefore, the signal did not contribute much to the overall variance seen in the data and was therefore neglected. To obtain the threshold estimates, the internal noise was re-generated 100 times, and the thresholds and the threshold error bars were computed in the same manner as the data thresholds (see Chapter 2, Methods section). In this section, we show the percent-greater functions for the data and the envelope-processing model as well as the threshold curves.

As we did for the neural responses, we defined threshold as the noise level where the signal could be detected for 75% of the stimulus presentations, through either an increase or a decrease in rate. The percent of stimulus presentations where there were more spikes for the signal-in-noise condition compared to the noise-alone condition for our two example units are shown in Figures 13 and 14, Panels A (S-90°, an unfavorable azimuth) and C (S+90°, a favorable azimuth). The translucent planes in these figures show the 25% (blue) and 75% (orange) criteria. The threshold in these figures is the highest noise level where the percent correct is either above the 75% plane or below the 25% plane. Panels B and D show corresponding responses for the envelope-processing model.

For the signal at -90° (Figure 13, Panel A for unit 22-0 and Figure 14, Panel A for unit 22-11) and the noise at positive azimuths, the signal tends to suppress the noise response. This suppression means that there are fewer spikes in the signal-plus-noise window than the noise-alone window, the signal is detected through a decrease in rate, and the percent-greater surface dips below the 25% plane (Panel A). For the signal and noise both on the ipsilateral side (Panel A, negative azimuths), neither the signal nor the noise excited the unit so that the percent greater stayed near 50% and the thresholds could not be determined. These conditions were rarely measured as there were few spikes, making it difficult to determine if a unit was still in contact with the electrode. For unit 22-0, the predicted suppression of the noise response (Figure 13, Panel B, solid arrow) is not as strong as the suppression seen in the data (Panel A, solid arrow), as can be observed by the shallower dip in the curve. Also, for this unit, the data shows that the signal can be detected through an increase in rate for the lowest noise levels (Panel A, dotted arrow). Such low level excitation is weaker in the model response, and not even present for some azimuths (Panel B, dotted arrow). The thresholds for the signal at -90° are plotted for unit 22-0 in Panel E and for the model in Panel F (solid black curves). For this unit and signal position, the threshold predictions appear to be similar for negative noise azimuths (although the thresholds were not measured for the noise at -90° or -54° for the data). However, for the positive azimuths, the model thresholds for 25% cannot be measured due to the weak suppression; these thresholds can be measured for the data where the suppression is stronger. For unit 22-11 and the signal at -90° (Figure 14, Panels A and B), the model predicts the unit's percent-greater curves fairly well. Here, the suppression for the S- 90° is more similar for the data and the model. Consequently, the thresholds for S- 90° are of similar magnitude and shape (Panels E compared to F, black curves); additionally, the thresholds cannot be measured for either the data or the model for any noise azimuth less than 45° .

For the signal at $+90^\circ$, shown in Panels C and D in Figures 13 and 14, both units detect the signal through an increase in rate, consistent with the model predictions (the percent-greater curve is above the 75% plane, solid arrow). The precise shape of the threshold curve, however, is more difficult to predict (red dotted curves in Panels E and F). For unit 22-0 (Figure 13), the shape of the threshold curves for the signal at $+90^\circ$ is

fairly well predicted by the model, but the amount of change in threshold is greater for the data than predicted. This difference in the amount of change of the thresholds is due to the fact that, for positive azimuths, the data thresholds are a little worse than the model thresholds, and for the negative azimuths, the data are a little better than predicted.

The model prediction for this unit's threshold curve when the signal was placed at 0° (Figure 13, Panel E and F, blue dashed line) is fairly good in that both the shape and the magnitude of the thresholds are well predicted. However, the relationship between the thresholds for $S0^\circ$ and $S90^\circ$ is not accurate because the model predicts that placing the azimuth at the favorable azimuth (90°) should improve the thresholds when the noise is also at a favorable azimuth. For the data, however, these thresholds are similar, and the only differences occur for azimuths near the midline. It seems that the actual unit is less sensitive to the actual position of the signal than the model predicts. Because these differences occur for the worst thresholds, these differences will have little effect on the population thresholds shown in Chapter 4.

In contrast, for 22-11 and the signal at $+90^\circ$ (Figure 14, Panels E, arrow, and F, red dotted curve), the data and the predictions of the thresholds differ for negative noise azimuths, but are similar for positive azimuths. There is an increase in the data thresholds for negative noise azimuths that is not predicted by the model. As discussed further below, the threshold increase for negative azimuths strongly depends on the exact model parameters, and so the threshold predictions could be improved by using a different set of parameter values. For unit 22-11 and the signal at 45° (magenta dash-dot line), the problems with the predictions are similar to those for the signal at 90° . The data shows a pronounced bounce in the thresholds for negative azimuths that is not as prominent in the model thresholds. However, the relationship between all of the curves for the noise at positive azimuths is similar for the data and the model, unlike the relationship between the $S0^\circ$ and $S90^\circ$ curves for unit 22-0.

Figure 15 shows selected thresholds for different units as a function of CF for the data, the modified cross-correlator model, and the envelope-processor model. In this figure, the thresholds for the $S+90^\circ$, $N-90^\circ$ thresholds (white squares) are compared to the $S-90^\circ$, $N+90^\circ$ thresholds (black circles). For all of these units, $+90^\circ$ is a favorable azimuth, and -90° is an unfavorable azimuth. These thresholds are particularly

interesting because of the cross-correlator model predictions. These signal and masker configurations give the largest separation of signal and masker so these configurations should give the best thresholds. Also, as discussed in Chapter 1, the overall correlation changes the most when the noise is excitatory and the signal suppresses the response, indicating that the best thresholds should occur for the S-90°, N+90° condition.

Somewhat unexpectedly, for the modified cross-correlator model, the signal was not detectable at any noise level for several units and conditions; however, as expected, when the signal could be detected in both cases, the S-90°, N+90° thresholds are slightly better than the S+90°, N-90°. The inability to detect the thresholds in the S-90°, N+90° condition can be attributed to the internal noise model; when the noise is removed, the S-90°, N+90° thresholds can usually be measured and are generally better than the reverse condition. However, the suppression of the noise response is not sufficient to detect the signal presence when noise is added. For the data, however, the reverse is true: the S+90°, N-90° thresholds are better or the same as the S-90°, N+90° thresholds.

The predictions of the envelope-processing model are somewhat better, but not very accurate. The S+90°, N-90° thresholds for the data show a fairly orderly improvement in threshold with CF; this orderly improvement is not seen in the envelope-processor predictions. Additionally, the thresholds can be measured in more cases than for the modified cross-correlator model, but not as many as in the actual data. Finally, for most, but not all, of these units, the model does predict that the S+90°, N-90° thresholds are best. As we will show in the next section, for the envelope-processing model, the S+90°, N-90° thresholds are particularly sensitive to the model parameters, especially those parameters related to the modulation filtering. Consequently, had we attempted to fit the parameters to give good threshold predictions, we might have improved the threshold predictions without significantly affecting the rate response predictions.

Overall, the thresholds were not as well predicted as the mean rate, probably because the unit parameters were fit to give good rate responses. However, as we will show in the next chapter, the model does predict the population thresholds, defined as the best threshold for all of the units at each signal and masker configuration, fairly accurately. In the next section, we discuss the threshold sensitivity to model parameters,

which should give some insight into why the threshold predictions are not very accurate and how these threshold predictions could be improved.

Effects of model parameters

In this section, we examine the effects of the model parameters on the single unit thresholds. Our goal for this section is to determine which parameter values give the lowest thresholds. We begin by varying a model unit's CF and best ITD while keeping the other parameters constant. Then we fix the CF and best ITD and vary the input-output functions one at a time to determine the effects of changing these functions.

The model parameters that have the largest effect on the model unit's threshold shape are the best ITD and the CF. Figure 16 shows the effect of varying the best ITD and CF for unit 22-0 for the signal placed at +90°; all of the other parameters for this unit remain fixed. In Panel A, we vary the CF while keeping the best ITD fixed at the measured best ITD (+290 μ sec, which corresponds to +90°) and show the responses. In order to see the effect of the CF on the rate-azimuth function, Panel C shows cosines with this equation

$$\cos(2\pi \text{ CF}(\Delta - \text{BITD}))$$

where Δ is the ITD associated with the noise azimuth, and the CF and BITD are the CF and best ITD of the model unit. (The transformation from ITD to azimuth is not completely linear, making the curves irregular.) The cosine is near 1 for favorable azimuths and near -1 for unfavorable ones. For changes in CF, the cosines predict the shapes of the curves fairly well: for suppressive noise azimuths, the thresholds are better (lower), and for excitatory azimuths, such as the one where the signal is placed, the thresholds are worse (higher). For the lower CFs, the best (lowest) thresholds occur for the masker in the ipsilateral hemifield (negative azimuths). For the highest CF of 1380 Hz (which was slightly higher than any CF in our population), the thresholds are sharply non-monotonic, with the thresholds getting worse as the noise moves to ipsilateral (and excitatory) azimuths.

Figure 16, Panel B shows the effect of changing the best ITD while fixing the CF. Varying the best ITD also changes the thresholds shapes, which are again similar to the shapes of the corresponding cosines (Panel D). In this case, because the best azimuth is

not fixed at the signal location ($+90^\circ$), the threshold curves are offset and scaled differently for the different best ITDs. For this unit, for example, having a best ITD of $-290 \mu\text{sec}$ (near -90° , blue dash-dot curves) places the *worst* azimuth near the signal location; these thresholds are the mirror images of the model thresholds for the signal at -90° for the fitted parameter set (Figure 13, Panel F, black solid curve).

Figure 17 shows that the best thresholds can be predicted based on the relationship of the locations of the signal and masker to the unit's best and worst azimuth. For these plots, we looked across all signal and masker configurations to determine the best overall threshold. We then looked at the position of the signal and masker at this best threshold location and compared that to the best and worst azimuths for the unit. Panel A shows a bubble plot of the best threshold signal location as a function of the unit's best azimuth. Panel B shows a histogram of the difference between these two azimuths. Because the difference is zero for the vast majority of the units, the best thresholds nearly always occur when the signal is placed at the unit's best azimuth.

The results for the masker location are less clear, as one would expect from the model thresholds in Figure 16. The best thresholds often occur when the masker is placed at the unit's worst threshold, but many occur for azimuths a little larger than the actual worst azimuth.

Because a unit's best ITD and CF have well-understood physiological meanings and the ranges of these values are well known, choosing ranges over which we varied these parameters was not difficult. The remaining parameters have less obvious physiological significance so that it is more difficult to know which values are appropriate. In order to ensure that the parameters that we use to study the parameter effects are realistic, we choose sets of parameters based on the variability of units in our sample.

In general, changes in parameters other than CF and best ITD cause smaller changes in the shapes of the threshold curves. Figure 18 shows the effects of different input-output functions for the boost. Panels A and B show the effect of varying the input-output function for the envelope processor for positive vector strengths. The vector strength is positive when the signal increases the overall rate so this input-output function only affects thresholds when the signal is at favorable azimuths. Panel A shows the

changes in threshold for the signal at $+90^\circ$, and Panel B shows the corresponding input-output functions. The black solid curves in Panels A and B show the actual model thresholds and input-output function for this unit, respectively, which gives fairly large changes in rate if the vector strength is large. For this curve, the rate does not change until the vector strength reaches the threshold value, vs_0 , of about 0.4, indicating that no boost will occur if there is a weak signal response. If vs_0 is lowered, but the slope is kept fairly constant (magenta solid line with circles), the unit's thresholds improve nearly uniformly for all of the masker azimuths. If vs_0 is greatly increased and followed by a steep slope (red dotted lines), the unit's thresholds do not change much from the base-line condition except for negative noise azimuths. These thresholds become worse, presumably because the boost was improving the signal response for these azimuths. If vs_0 is lowered, but the overall boost is small (dashed green line with x's), the unit's thresholds improve dramatically for a few negative noise azimuths and then get worse, giving a "bounce" in the thresholds, again presumably because the signal needs an increased rate to give the improved thresholds. Finally, if the boost is negative, which would decrease the signal response, the thresholds become uniformly worse.

The input-output functions for the negative vector strengths (Figure 18, Panel D) have much less effect on the thresholds (Figure 18, Panel C). Again the solid black curve represents the actual fit to the data. Because the negative vector strengths occur when the signal suppresses the noise response, these input-output functions affect the thresholds for the signal at unfavorable azimuths, -90° in this case. When vs_0 is near 0 (magenta solid line with circles) or the input-output function has a large final value (red dotted line with x's), the thresholds can be measured for the S- 90° , N 90° and S- 90° , N 72° conditions. It is possible that more drastic changes in the input-output function for negative vector strengths would improve these thresholds further.

We mentioned above that there seemed to be a value of the correlation ρ_0 below which the boost did not occur. We imagined that inputs to the mechanism that provided the boost was inactivated for small correlations. Consequently we changed ρ_0 , the value of correlation below which there is no boost, in Panels E and F. If ρ_0 is near zero, then the boost is never inactivated by low correlation values; if ρ_0 is near 1, then the boost is always inactivated. Consequently, ρ_0 shows the overall effect of the boost for this unit.

For this set of parameters, the boost controls whether there is a “bounce” in the thresholds: if the boost is inactivated for some correlation values, the thresholds for the noise at negative azimuths do not improve monotonically as the noise is moved away from the signal. If the boost is not inactivated, the threshold curve is monotonic.

Figure 19 shows the effects of the rate input-output function (Panels A and B) and the correlation input-output functions (Panels C and D). For the rate input-output functions (Panels A and B), the thresholds are poor when the second line in the input-output function has zero slope (cyan dash-dot curve with circles) or negative slope (magenta solid curve with circles; green dashed line with x's) near the signal-alone rate (dotted lines). The best threshold occurs for a rate input-output function with a steep slope for all input rates (red dotted line with x's). There is a simple explanation for this effect. The signal-alone rate is usually around 8 spikes/stimulus, which is marked by the dotted black lines in Panel B. As long as the input-output functions have about the same slope near this point, then the noise-alone and signal-plus-noise rates increase with increasing noise level at the same rate (black solid curve; red dotted curve with x's; blue dash dot curve with x's). However, for the other curves (green dashed curve with x's; magenta solid curve with circles), the noise-alone rate approaches the signal-plus-noise rate more quickly, making the thresholds worse since it is easier to mask the signal.

The correlation input-output functions show similar results for similar reasons. In this case, the input-output functions are rectified parabolas. Consequently, the slope of the input-output function changes with the input correlation. The best thresholds occur for parabolas that increase quickly with correlation (solid magenta curves with circles, dash-dot blue lines with x's). When the parabola is steep, the signal-plus-noise rate actually increases faster than the noise-alone rate with increases in noise level, and the signal is more difficult to mask, which improves the thresholds.

The model parameters can dramatically change the threshold curves, and changing different parameter sets can give similar effects. The CF and the best ITD change the shapes of the curves as well as their overall magnitude; in general, the remaining parameters cause less dramatic changes in the shapes of their curves, but instead shift portions of the threshold curves up or down. From this analysis, it appears that the most sensitive unit for a given signal and masker configuration will have its best

ITD at the signal azimuth and a CF so that the worst ITD is either at the masker azimuth or just on the side farthest from the signal. Furthermore, a large boost that affects even small positive vector strengths and all correlation values, as well as expansive rate and correlation input-output functions will improve the unit thresholds.

As noted earlier, the thresholds for large separations of the signal and masker are not always well predicted (see Figure 15). However, as shown in this section, these thresholds are particularly sensitive to changes in the signal-induced changes in rate (Figure 18).

Discussion

In this chapter, we developed a model of ITD-sensitive, low-frequency units in the IC. The purpose of this model was to determine what processing was necessary to predict the single unit responses and thresholds. This model, which was developed by comparing the model rate responses to the data, has simple processing paths and fairly accurately predicts the rate responses of all the units in our sample. In order to predict the responses, we needed a mechanism that generated rate-level functions and rate-ITD functions independently, and a mechanism that changed the overall rate based on the strength of the signal compared to the noise. That we would need a rate-ITD processing path is not surprising, and we borrowed a well-known model similar to that of Jeffress (1948) and Colburn (1973, 1977a, b) to generate ITD sensitivity. We found, however, that the cross-correlation alone was unable to explain the noise-alone response as a function of noise azimuth and noise level. In order to account for this response, we employed two independent processing pathways, one for computing the normalized correlation of the delayed outputs of the ANFs and one for computing the average output rate of the two ANFs. These inputs gave reliable estimates of the noise level and azimuth, which we were then able to combine to create the entire noise-alone response matrix.

That the responses were independent would suggest that the mechanism for ITD-sensitivity is independent of the overall level. Goldberg and Brown (1969) report that the best and worst ITDs for units in the MSO do not change with variations in the overall level; the one example unit appears to have ITD-sensitivity independent of the overall

level, indicating the MSO responses may be independent with respect to level and ITD. Due to the paucity of reports on MSO responses, we hesitate to speculate where or how the independent processing of level and azimuth occur.

The third processing path added sensitivity to envelope modulation. This addition to the modified cross-correlator model was necessary because we observed that the signal-alone response was different from the noise-alone response, even when they were matched in level and azimuth. The modified cross-correlator model alone could not replicate this response. Several researchers (e.g. Langner and Schreiner, 1988; Krishna and Semple, 2000) have shown that many units in the IC have an enhanced response to modulation rates less than 100 Hz, and the best modulation rates are often near 40 Hz, which is the modulation rate of the signal used here. Consequently, we used the 40-Hz Fourier component of the response after the cross-correlation to determine the amount by which the overall rate should be changed. It seems unlikely that the actual units are changing the rate based on the 40-Hz component alone. A more plausible explanation is that there is a combination of excitation and suppression timed in such a way as to enhance the response to certain modulation frequencies and suppress the response to others. Figure 20 shows how the timed excitation and suppression could give modulation sensitivity. Panel A shows a result similar to those shown in precedence effect studies like those of Yin (1994). This panel shows the response to a leading binaural click (dash-dot line) and a lagging click (solid line). When the lagging click closely follows the leading click, the response to the lagging click is suppressed. (For the smallest lags, the response is recorded for the leading click.) From these extracellular recordings, it cannot be ascertained whether the suppression occurs due to local inhibition, inhibition in lower auditory nuclei, or some other suppressive mechanism. However, Panel C shows how an inhibitory mechanism could suppress the response to the lagging click: the dotted line shows an alpha-function is meant to mimic the response of a fast excitatory post-synaptic potential (EPSP) while the dash-dot line shows a slower inhibitory post-synaptic potential (IPSP). The solid line shows the sum of the two curves, which results in excitation followed by inhibition. Such timed excitation and inhibition could act as a filter, essentially differentiating the inputs to a unit. The first click would therefore cause excitation and long-lasting inhibition so that the leading click would suppress the

response to lagging click. Panel B shows the rate modulation transfer function (MTF) for the unit in Panel A. Here the rate changes with the modulation frequency of sinusoidally modulated noise bursts. This unit prefers modulation frequencies near 30 Hz. Panel D shows the frequency response of the filter shown in Panel C. The frequency response is similar to the modulation transfer function, indicating that this filter could provide the sensitivity to modulation rate seen by some of these units. We note that neural inhibition is not absolutely necessary to create bandpass modulation transfer functions. Instead, neural adaptation could also be responsible; in this case, the neuron would not be ready to immediately re-fire causes a similar kind of lagging suppression.

We showed here that the boost appears to occur after the cross-correlator because the amount of rate change required strongly depends on the signal azimuth. That the boost occurs after the cross-correlation suggests that the underlying neural mechanism could occur at or beyond the MSO. As we discussed above, it is possible for the boost to occur due to excitation followed by inhibition, giving a kind of temporal differentiator. This pattern of inputs has been shown to occur locally in IC units (Kuwada et al., 1997). Additionally, through the use of a different style of computational model of IC units, Cai et al. (1998) have shown that delayed inhibition is sufficient to describe a wide variety of IC responses and postulated that this inhibition may come from the dorsal nucleus of the lateral lemniscus (DNLL).

While the idea that single IC units have bandpass MTFs is by no means novel, the fact that these MTFs could play a role in binaural signal detection seems to be. The envelope processing improves the thresholds for the cases when the signal is at a favorable azimuth and the noise is at an unfavorable one. As we showed, there is much less need for a boost for the case when the signal is at an unfavorable azimuth. Because of this advantage of placing the signal at favorable azimuths, these thresholds become better than the ones for the signal at unfavorable ones, which we see both in our data and our model results. Without this signal enhancement at favorable azimuths (the modified cross-correlator model) or without a difference in the modulation of the signal and noise (as for tones in noise, see previous chapter and e.g. Palmer et al, 1997b), the thresholds are better when the signal suppresses the noise response, consistent with the idea that $N0S\pi$ thresholds are better than $N\pi S0$. This well-known result does not seem to hold for

these units and stimuli, which are broadband, have different modulation rates for signal and noise, and are placed at actual locations in space, and are therefore somewhat more natural conditions than the ideal $N0S\pi$ and $N\pi S0$ conditions usually tested.

We chose a 40-Hz modulation rate for our signal because many IC units show an enhanced response for this modulation rate. We used the chirp trains as our search stimulus, practically guaranteeing that these were the units we found. Other modulation rates may have been preferred by other units, however, so it is possible that, within limits, similar results will be found for signals with different modulation rates. Any other processing that could successfully separate the signal response from the noise response is likely to produce similar results. Few unrelated choices exist, however, because the primary difference between the broadband signal and the broadband noise is in their temporal characteristics.

Additional insight into the potential underlying neural mechanisms came from varying the model parameters. The most dramatic change in the thresholds came from varying the CF and the best ITD because these parameters change which azimuths are favorable. As described in Ch. 2, the most effective masking azimuth is often the best azimuth, and our analysis here demonstrates this idea as well. The shapes of the model threshold curves were largely determined by the CF and best ITD so that when the best and worst ITDs were on opposite sides of the head, the thresholds curves were nearly monotonic, with the best thresholds on the contralateral side and the worst thresholds on the ipsilateral side. In contrast, units that had the best and/or worst ITDs inside the physiological range had non-monotonic threshold curves. It has been shown recently that units in the IC with high CFs have best ITDs near zero and units with low CFs have large best ITDs (McAlpine et al., 2001). This result explains why we generally see a continuum between two types of threshold curves in our population: the nearly monotonic ones and then ones that peak near 0° for higher CF units. We will show in the next chapter that a population of these two types of threshold curves can predict the psychophysical thresholds in humans, a result which may have implications on the distribution of best ITD and CF in humans.

The next set of parameters, the ones related to the boost, also affected whether the threshold curves would be monotonic. In general, an increase in the boost made the

signal more easily detectable and therefore improved thresholds. The observations that the boost only occurred if the inputs were above some threshold correlation value demonstrated this idea effectively (Figure 18, Panels E and F). If the noise was placed at an unfavorable azimuth and was able to reduce the correlation below some correlation threshold, then the boost did not occur. This effect created a “bounce” in thresholds, which was also seen in the data. Physiologically, such an effect might occur if the input to the hypothesized delayed inhibition were sub-threshold. In this case, the temporal differentiator would not be activated, removing the high-pass nature of the MTF and increasing the response to the noise. In this case, the signal and noise responses would be more similar, making the thresholds worse and causing thresholds for the noise at negative azimuths to become worse. This idea is consistent with the fact that some neurons show lowpass MTFs at low SPLs and highpass MTFs at higher SPLs (Rees and Møller, 1987).

The input-output functions for the rate and correlation also affected thresholds. Essentially, the more expansive the input-output curves, the better the thresholds because for expansive functions, the noise response did not increase as much as the signal-plus-noise response with increasing noise levels.

Overall, we created a model structure that was essentially derived from the data. This model, a type of “inverse model” as described by Zweig (1991), allowed us to understand the processing that was required to predict the individual unit responses. The model as we have created it cannot reliably predict novel stimuli because such stimuli will likely require additions to the model that we did not foresee. However, since the purpose of the model is to understand the neural processing that resulted in the known responses, we feel that there is merit to the model despite this drawback. Once the actual neuronal mechanisms have been more thoroughly explored, a predictive and physiologically viable model might be feasible. Furthermore, because an auditory nerve model similar to the one used here also exists for humans and hearing impairment, one could use this model to begin exploring the differences between cat and human responses and the changes that may occur with hearing impairment. In the end, the processing required to predict the neural responses—cross-correlation, rate-level functions with varying shapes, and sensitivity to the signal envelope—have all been observed in single

IC units by other researchers. However, the finding that these neural mechanisms interact to play primary roles in low-frequency signal detection is novel. These three processing paths together were able to explain a large amount of data, as well as explain a difference between the current work and previous physiological data, where the best thresholds for tones in noise occurred when the signal was given an unfavorable interaural phase. The next chapter shows how the population responses of both the actual units and the model units compare to psychophysical results.

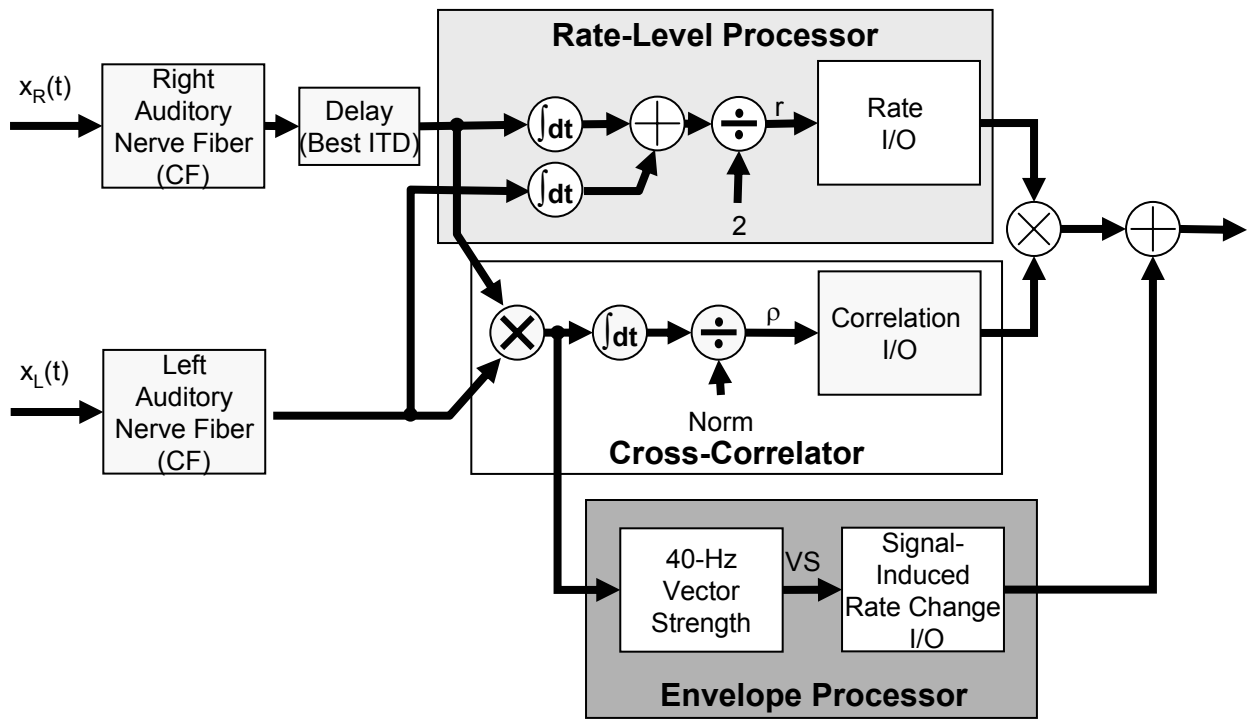


Figure 1: Model block diagram. The white block indicates the cross-correlator, the light gray block indicates the rate-level processor, and the dark gray box indicates the envelope processor. The white and light gray paths together constitute the modified cross-correlator model while all of the blocks together constitute the envelope-processing model. The final blocks for each pathway show the input-output functions. Besides the CF and best ITD, most of the model parameters are incorporated into these input-output functions. In some cases, if the output correlation ρ is below a threshold value ρ_0 , then the envelope processor will be disabled (not shown). For the threshold predictions, an internal noise sample is added to the output of the model (not shown).

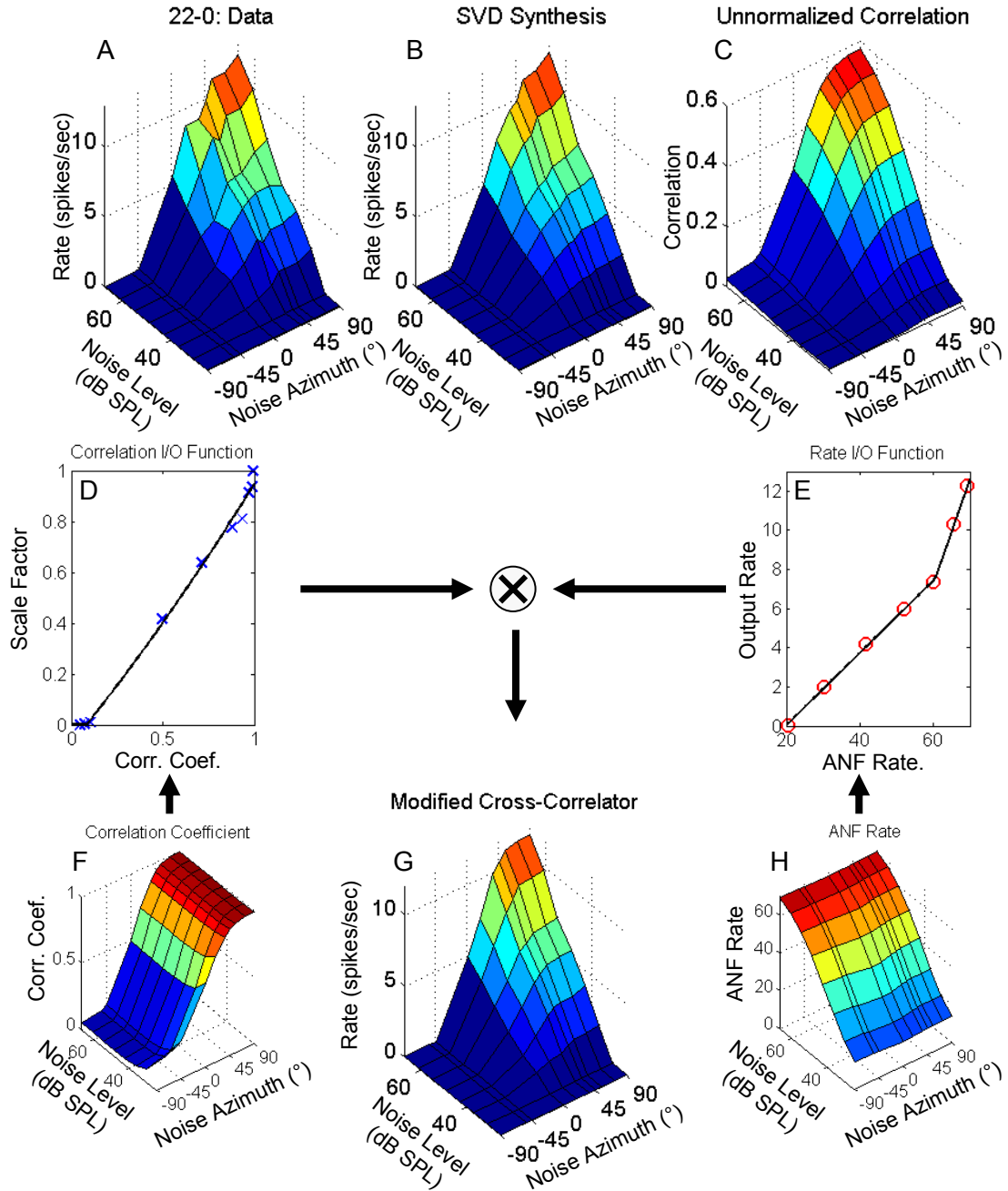


Figure 2: Noise-alone response for unit 22-0 and model. (A) Noise-alone rate response as a function of noise level and azimuth for unit 22-0. (B) The reconstruction of the rate response from the SVD analysis. (C) The unnormalized interaural correlation as a function of noise level and azimuth. Because there is no normalization for overall energy, this variable depends on both noise level and noise azimuth. Since both the rate-level function and rate-azimuth functions are monotonic, the dependence on noise and level resemble the data's dependence on these variables. (D) Correlation input-output function fit with a parabola and a threshold. (E) Average rate input-output function fit with two connected lines. (F) Correlation coefficient as a function of noise level and azimuth. The normalization causes this variable to be nearly independent of noise level. (G) The modified cross-correlator model response. (H) Rate averaged across both auditory nerve fibers as a function of noise azimuth and level. This variable is nearly independent of noise azimuth.

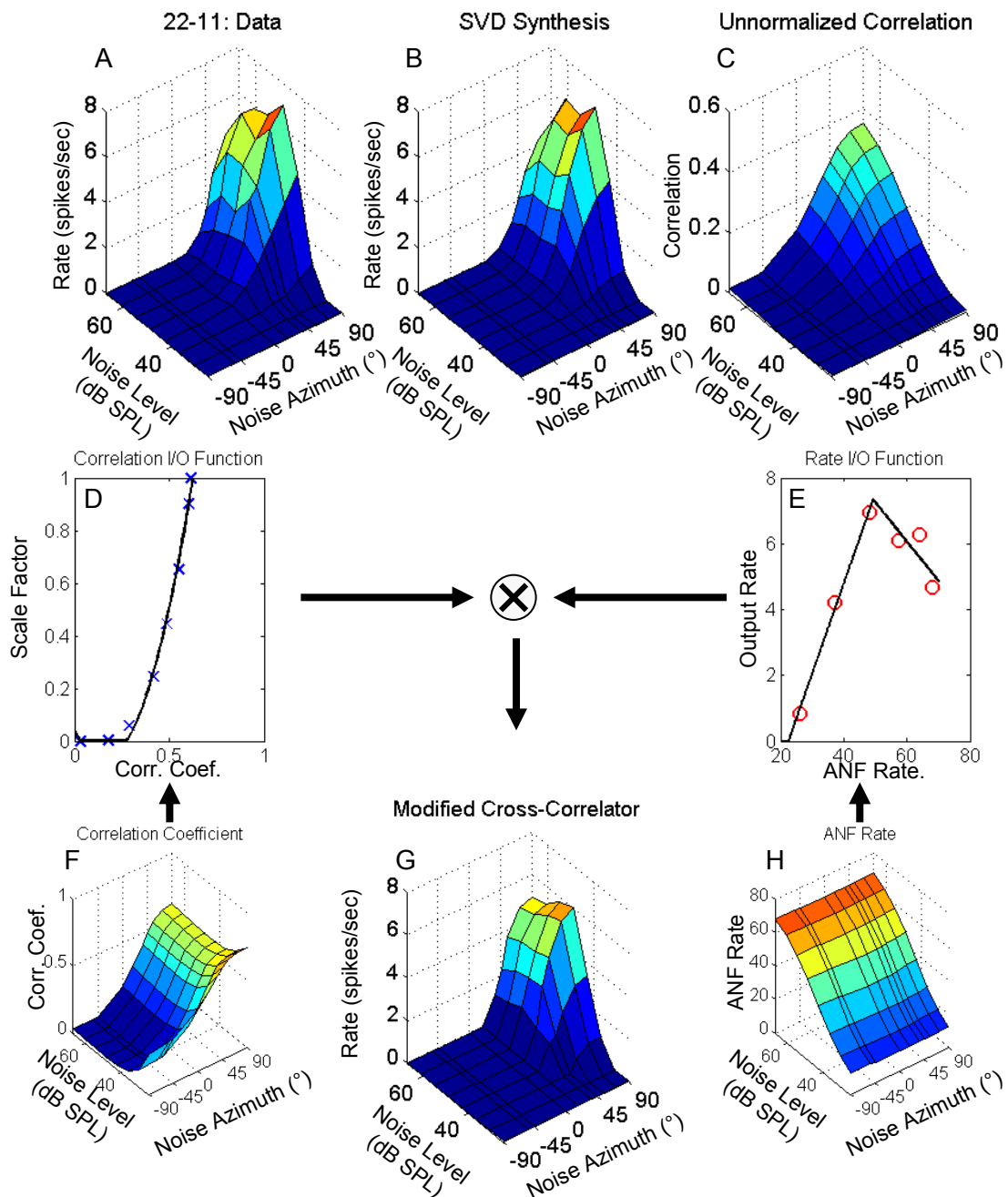


Figure 3: Noise-alone response for unit 22-11 and model. Same format as Figure 2. Note that the non-monotonic rate-level function for the data (A,E) cannot be replicated by the unnormalized correlation (C).

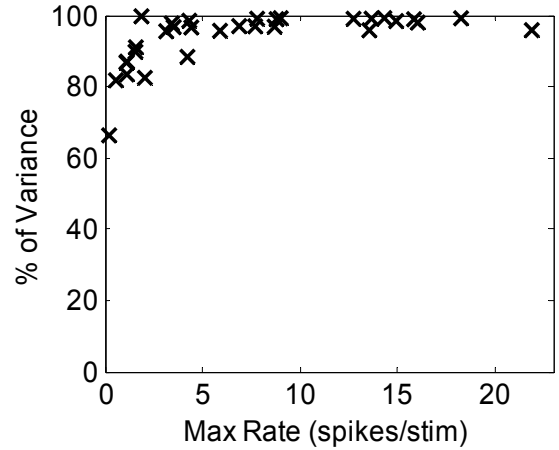
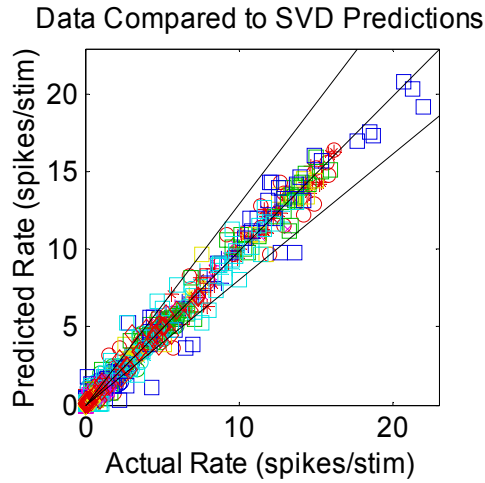


Figure 4: Left: Actual rate of for the individual units (different markers and colors) compared to the rate by assuming that the rate and level affect the overall response independently using SVD analysis. For the data, the variance in the noise response is generally proportional to the square of the overall rate. The proportionality constant is usually less than, but near 1. To give an estimate of the standard deviation for the rate, we calculated the proportionality constant that minimizes the error for each unit individually. The outer lines indicate ± 1 standard deviation using the average of all the proportionality constants. Right: The percent of variance accounted for by the SVD predictions as a function of the maximum noise-alone response. For units with large noise-alone responses, nearly all of the variance in the data could be accounted for with the SVD analysis.

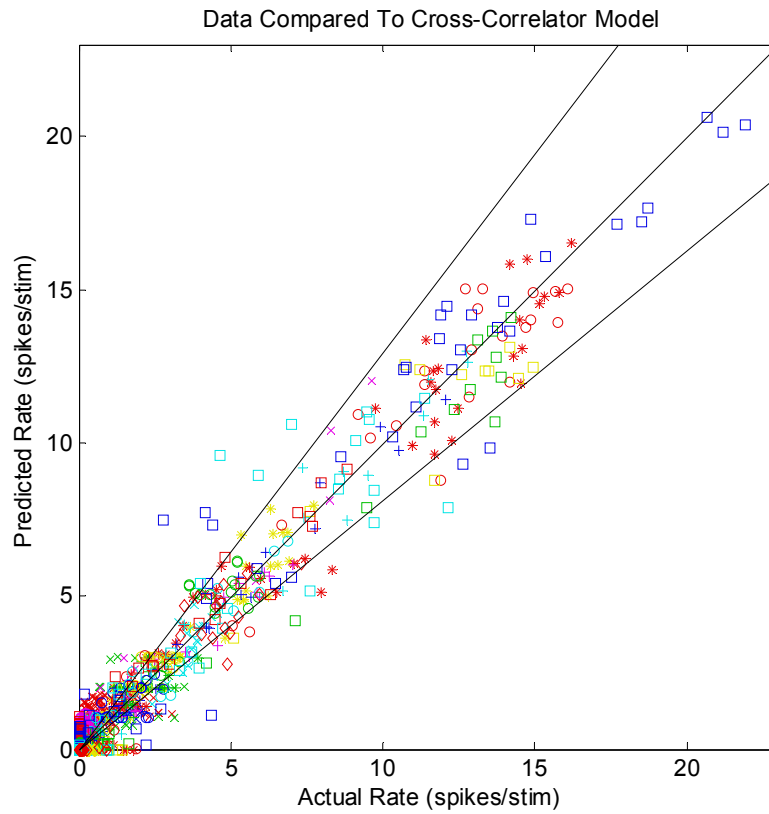


Figure 5: Same format as Figure 4 for the modified cross-correlator model. The assumptions of the model have made the predictions worse, but are generally still accurate.

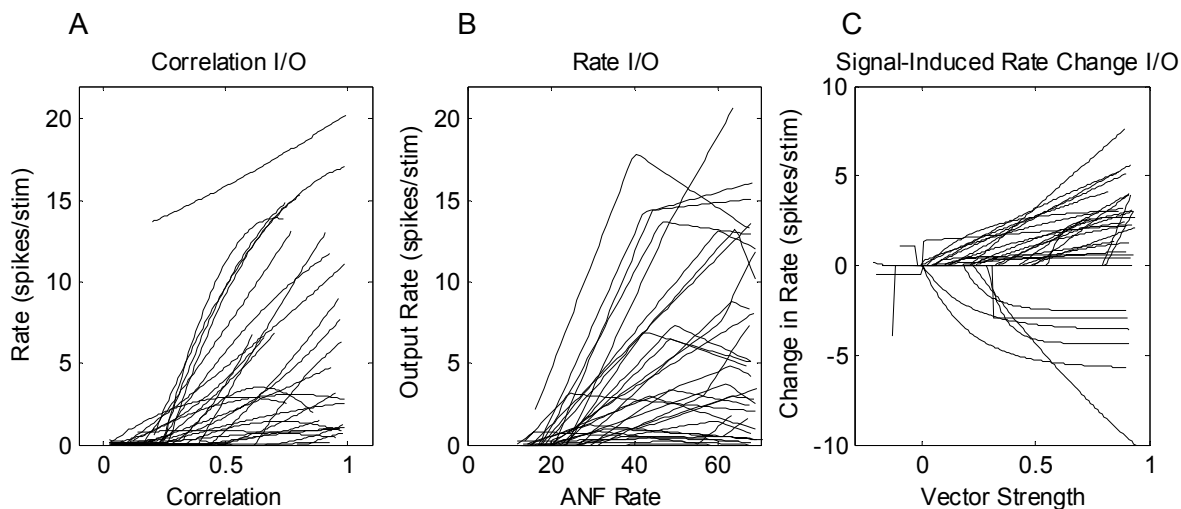


Figure 6: Input-output functions for all of the units in the population. (A) Parabola-with-threshold fit to correlation input-output functions. The functions are scaled by the maximum output rate. (B) Two-line-with-threshold fit to rate input-output functions. (C) Exponential decay fit to change-in-rate input-output functions.

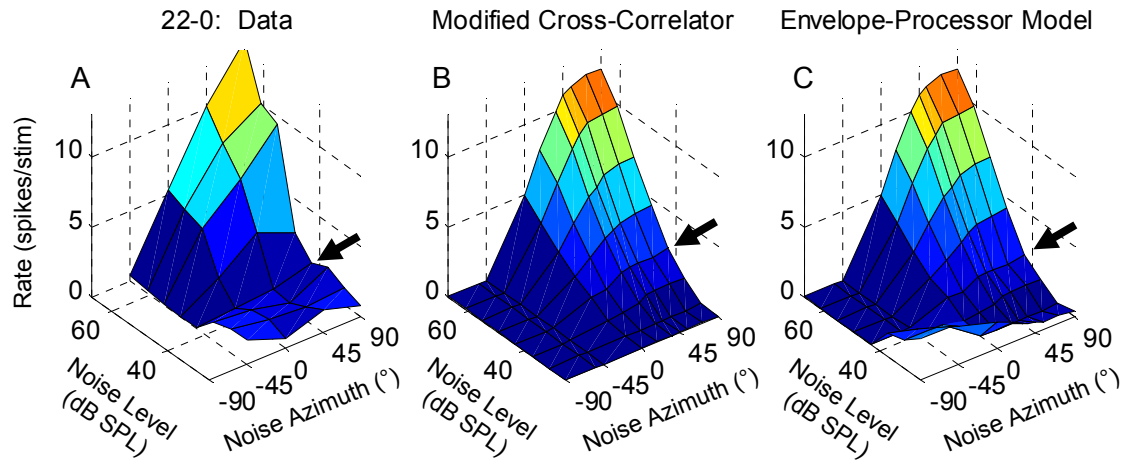


Figure 7: (A) Individual unit response for 22-0 for the signal at -90° . The arrows in A-C point to the suppressive effect of the signal. The modified cross-correlator model (B) and the envelope-processor model (C) give similar predictions; both predict a slight decrease in the rate for the noise at $+90^\circ$. The suppression is not as large as that seen in the data. The envelope-processor model does predict the weak excitatory response at low levels for some noise azimuths similar to the data.

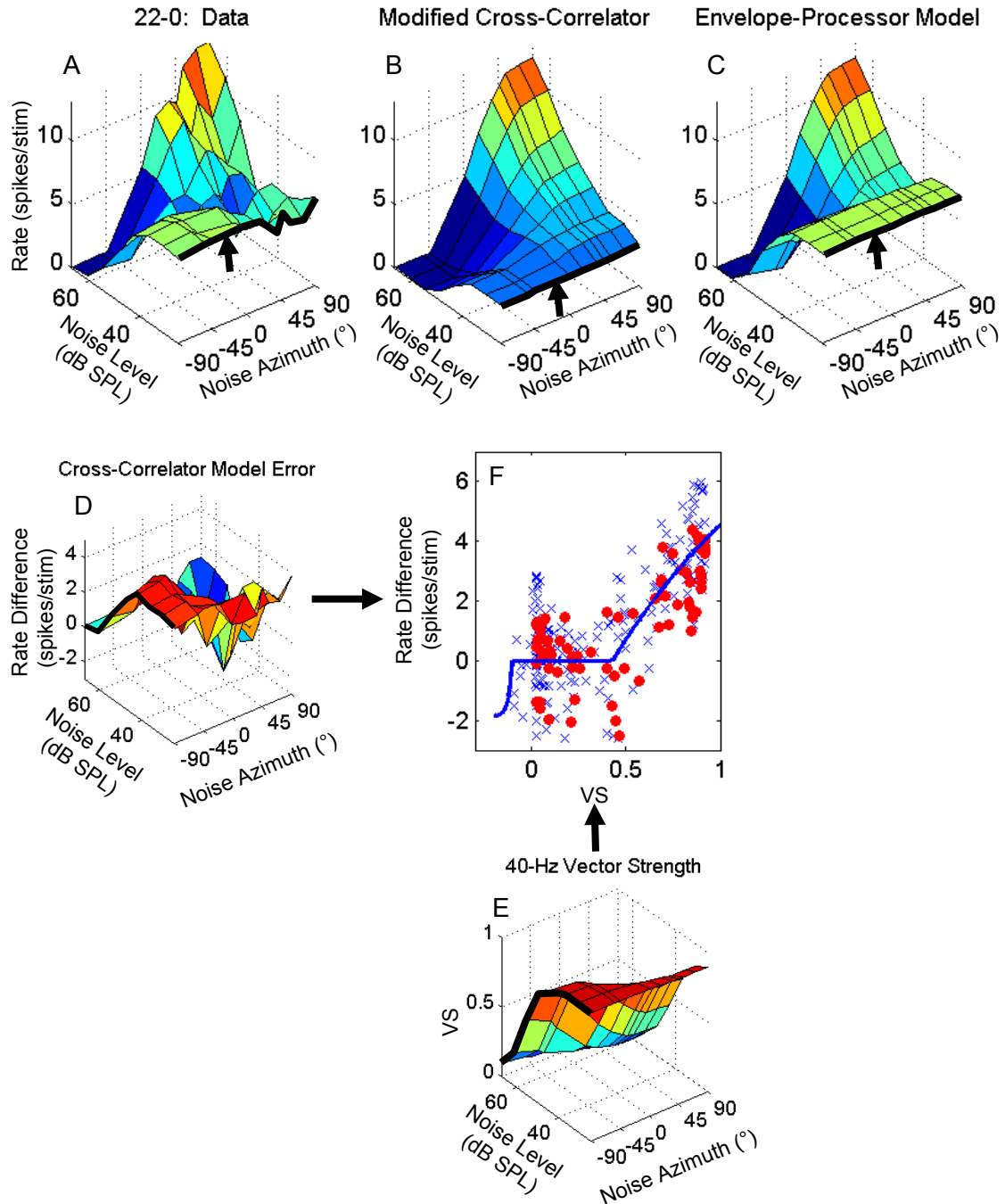


Figure 8: Signal-plus-noise response for unit 22-0 and signal at +90° as a function of noise level and azimuth.. (A) Actual unit response. Signal-alone response can be seen when the noise level is low (edge indicated by arrow). (B) Modified cross-correlator model response. Signal-alone response is underestimated. (C) Envelope processor model response. Signal-alone response is improved. (D) Difference between actual response and modified cross-correlator response. Largest difference occurs for low noise levels. Outlined edge for noise at -90° emphasizes the change with noise level. (E) Vector strength (VS) of the response after the multiplier for the signal-in-noise response. This function has a somewhat similar shape to D. (F) Required difference as a function of VS difference. Blue x's show data for all noise azimuths, red dots are for the data shown in this figure, S+90°. The VS is negative when the signal suppresses the noise response. Blue line shows fit described in text.

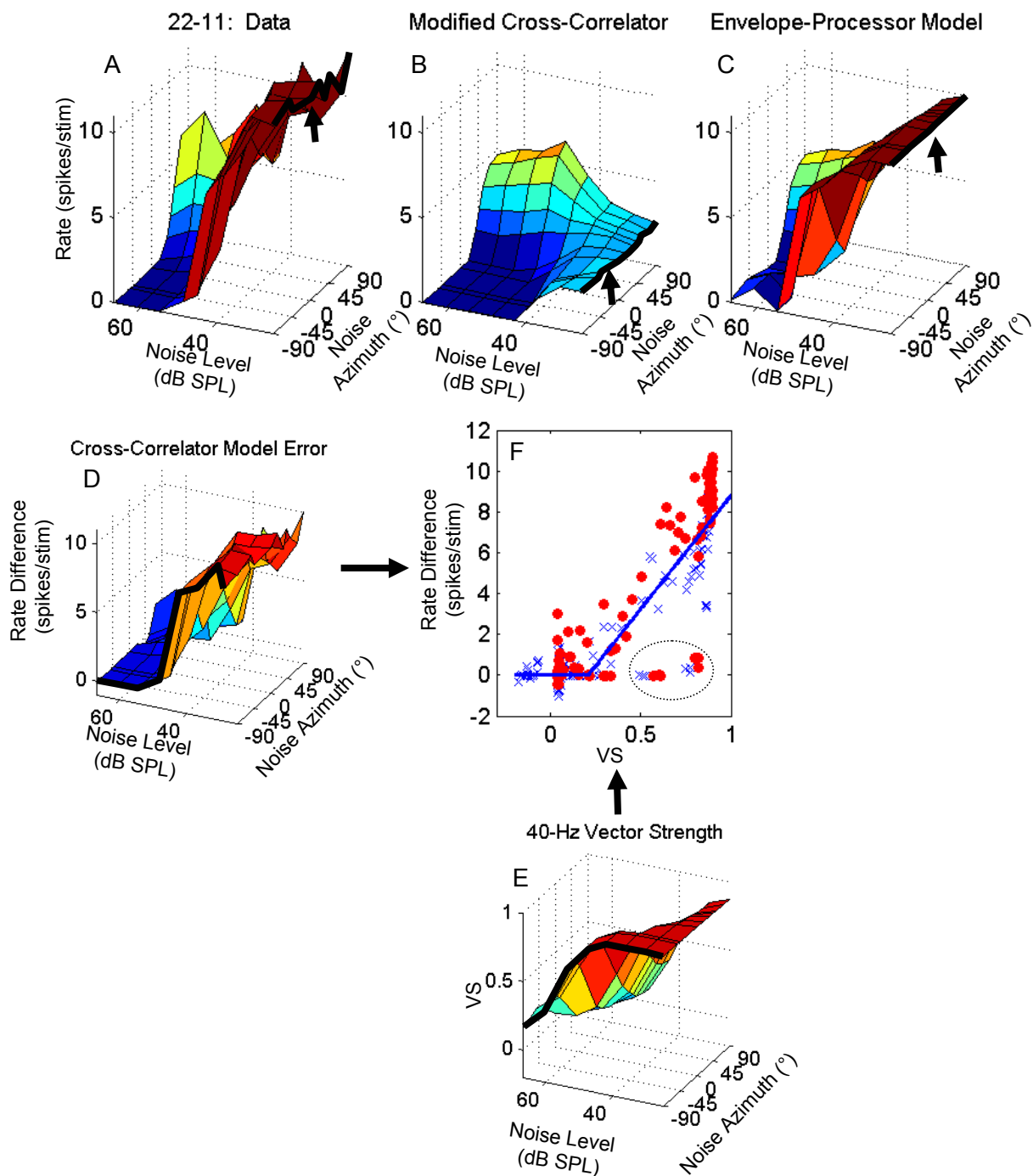


Figure 9: Same as Figure 8, but for unit 22-11. The points inside the oval in F are the outliers for which the correlation was below the threshold for the signal-induced rate change (see text).

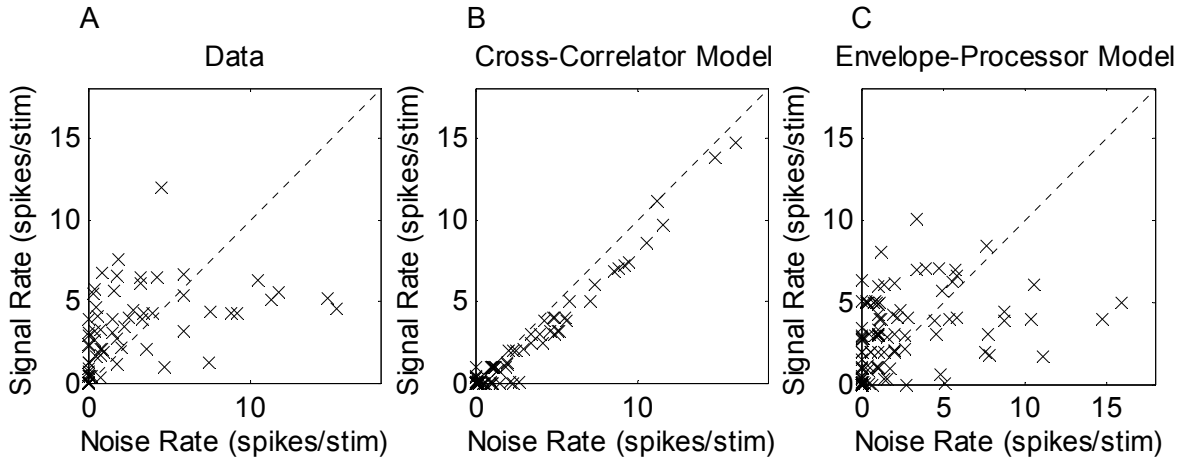


Figure 10: Signal-alone rate compared to noise-alone rate response. For the data (A), the signal-alone rate is usually higher than the noise-alone rate when they are both at the same level and azimuth. For the modified cross-correlator model (B), the signal rate and noise rate are similar, except that the signal rate is slightly smaller than the noise rate. For the envelope-processor model (C), the relationship between the signal and noise rates is similar to the data.

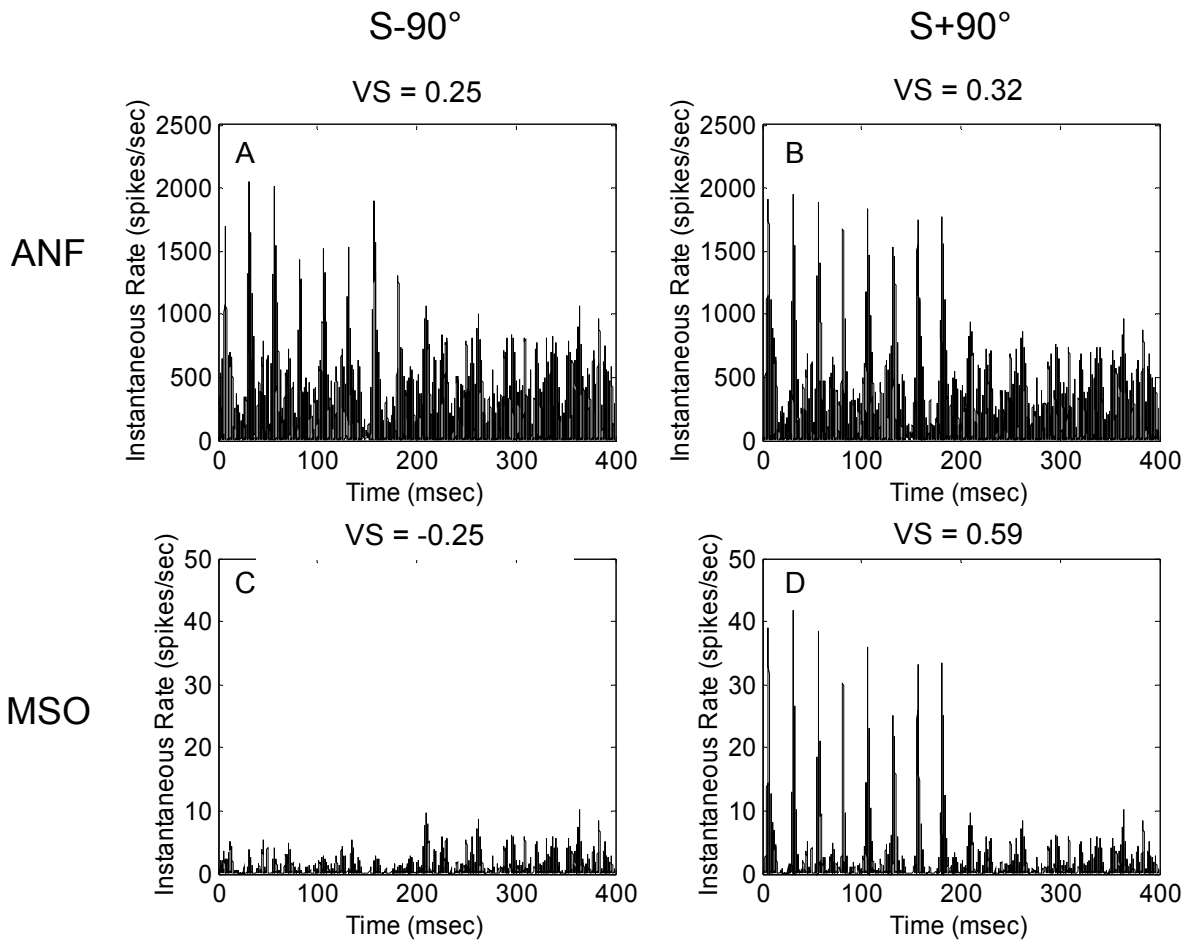


Figure 11: Model auditory nerve fiber response (A,B) compared to output of the multiplier, which is something like the output of an MSO cell (C,D) for signal at -90° (A,C) and signal at $+90^\circ$ (B,D). 0- 200 msec show the signal-plus-noise response while the 200-400 msec shows the noise-alone response. The corresponding difference in vector strengths between the signal-plus-noise and noise-alone conditions are shown above the panels. While the auditory nerve fiber responses and vector strengths do not change much with large changes in signal azimuth, the MSO response shows a large change with the signal azimuth. Additionally, the vector strength changed sign because for -90° , the signal suppresses the noise response, giving excitation in the valleys between the response to the chirps.

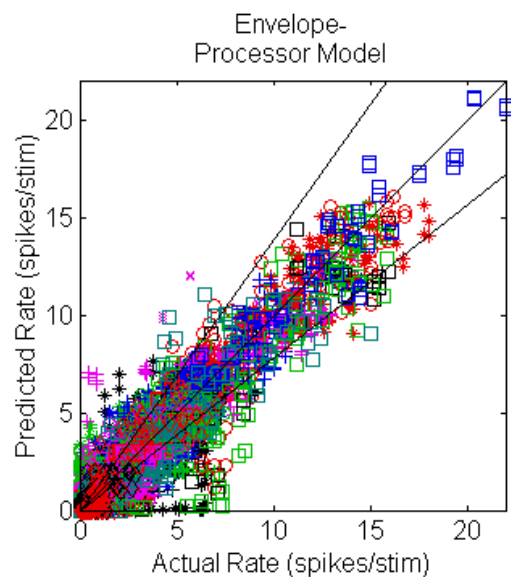
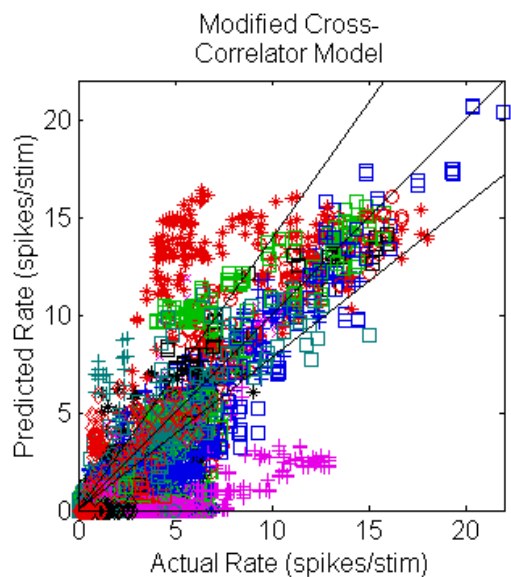


Figure 12: Signal-plus-noise rate for modified cross-correlator model and envelope processor model compared to data. The envelope-processor model predicts the data more reliably than the modified cross-correlator model. Lines same as Figure 4.

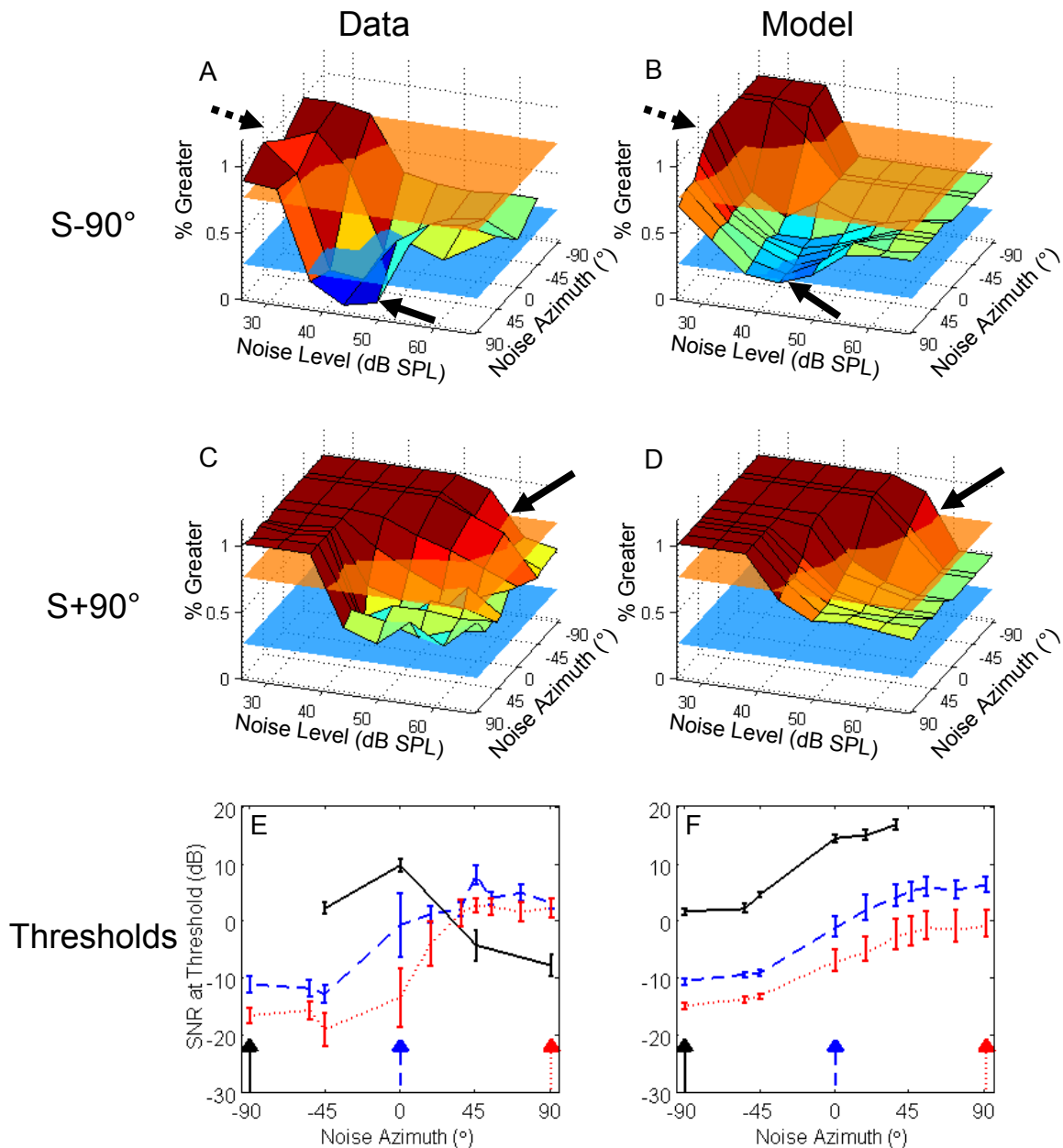


Figure 13: Data and model thresholds for unit 22-0. Panels A-D show the percent of times there were more spikes in the signal-plus-noise window than the noise-alone window. Plots are in a different orientation (positive azimuths towards the front) than rate response plots so that the important features can be seen. Horizontal planes show 25% and 75% criteria. The signal can be detected when the % greater is either above the 75% plane or below the 25% plane. If the % greater is above the 75% plane, the signal is detected through an increase in rate, and if the % greater is below the 25% plane, the signal is detected through a decrease in rate. (A) 22-0 response for signal at -90° . Solid arrow points to suppression by the signal, dotted arrow points to excitation by the signal for low noise levels. (B) Model response for signal at -90° . Arrow points to suppression by the signal, which is too small to reach threshold. (C) 22-0's response for signal at $+90^\circ$. Arrow points to the threshold line. (D) Model response for signal at $+90^\circ$. Arrow points to threshold line. (E) Thresholds as a function of noise azimuth for several signal azimuths for the data. (F) Same as E for the model.

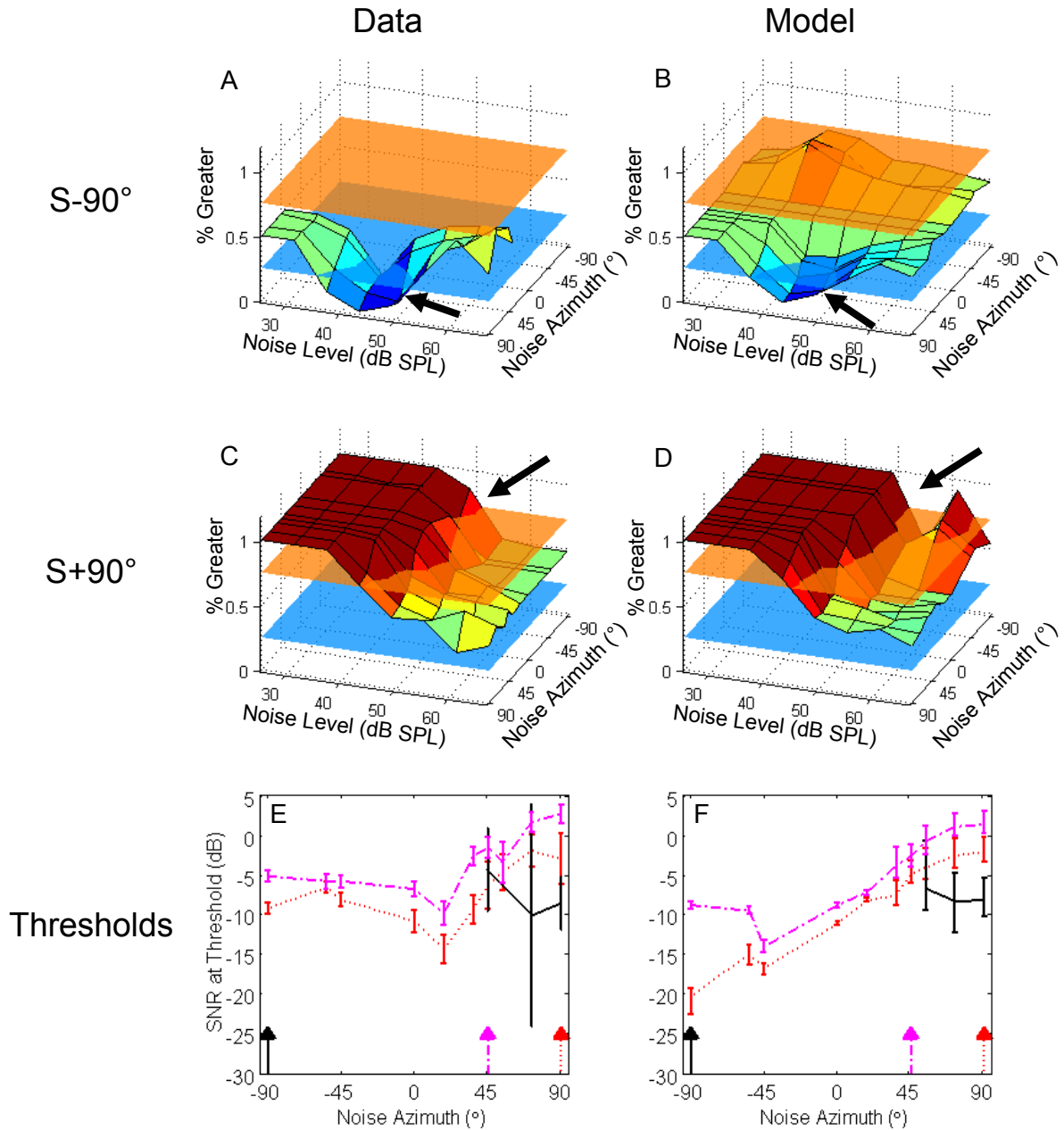


Figure 14: Data and model thresholds for unit 22-11. Same format as Figure 13. For B, the signal does suppress the noise response enough to reach threshold. For D, a non-monotonicity in the %-correct curve at -90° makes the thresholds better than the actual thresholds.

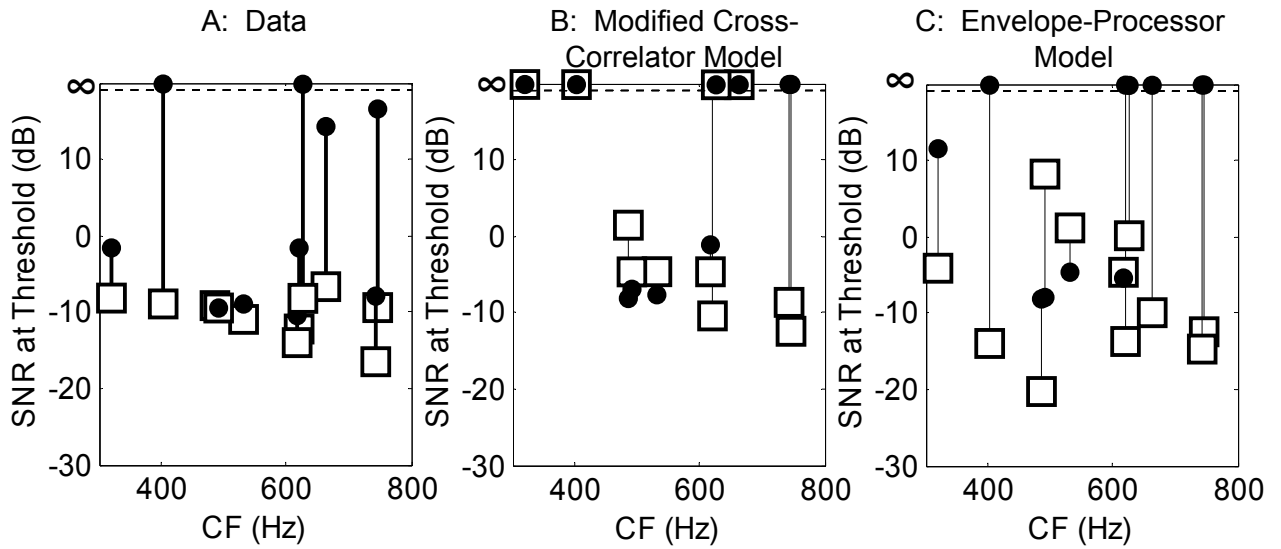


Figure 15: Thresholds for signal at $+90^\circ$ and noise at -90° (white squares) compared to thresholds for signal at -90° and noise at $+90^\circ$ (black circles) for data (A), modified cross-correlator model (B), and envelope-processor model (C). Infinite thresholds indicate that the signal could not be detected at any noise level.

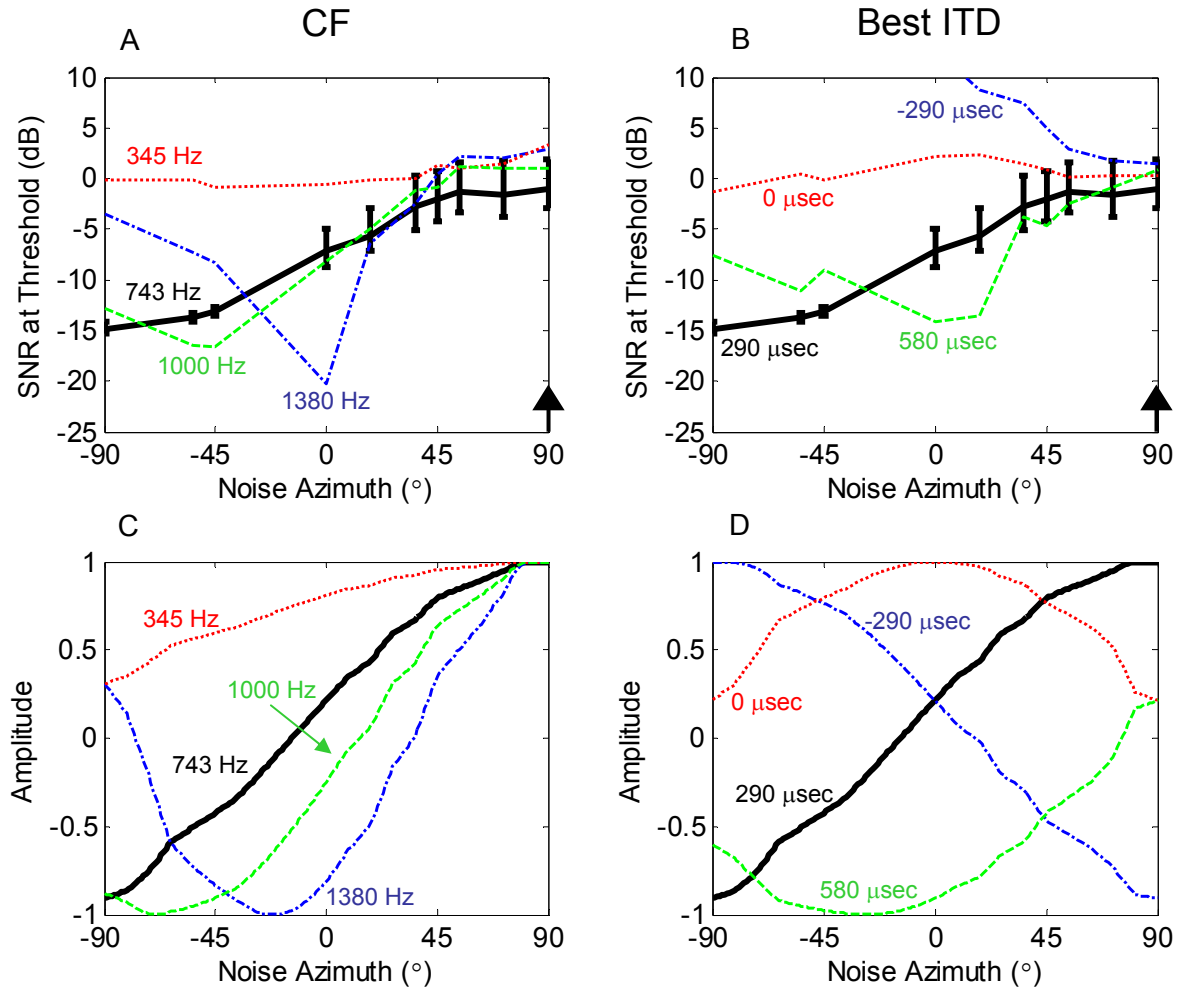


Figure 16: Model thresholds for 22-0 for different CFs (A) and best ITDs (B) for the signal at +90°. All other parameters were fixed to original values. Original values indicated by thick black lines. Panels C and D show $\cos(2\pi \cdot \text{CF}(\Delta\text{-BITD}))$, which shows an approximation of how the normalized rate-azimuth would change with changes in CF (C) and best ITD (D).

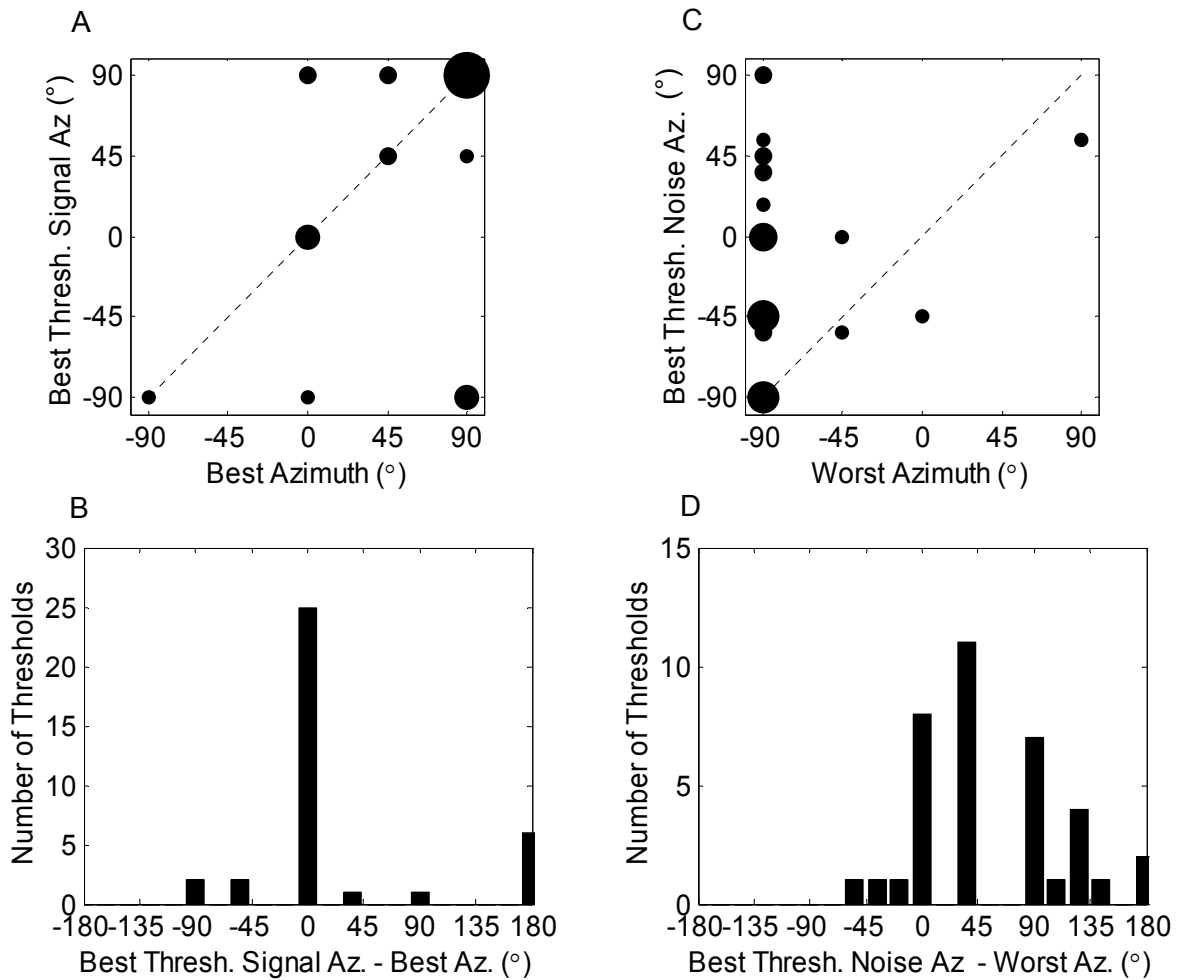


Figure 17: Best thresholds for the envelope-processor model usually occur for the signal at the best azimuth and near the worst azimuth. (A) The signal azimuth for the unit's best overall threshold as a function of the model unit's best azimuth. The size of the dot is proportional to the number of points at that location. The best thresholds usually occur when the signal is placed at the unit's best azimuth. (B) Histogram of the difference in the best azimuth and the best-threshold signal azimuth. Nearly all the best thresholds occur when the signal is placed at the best azimuth. There are more thresholds than units in the population because sometime there were multiple configurations that gave the same best threshold. (C,D) Similar to A and B for the masker azimuth at the best threshold and the worst azimuth. The histogram shows that the best thresholds do not necessarily occur when the masker is at the worst azimuth.

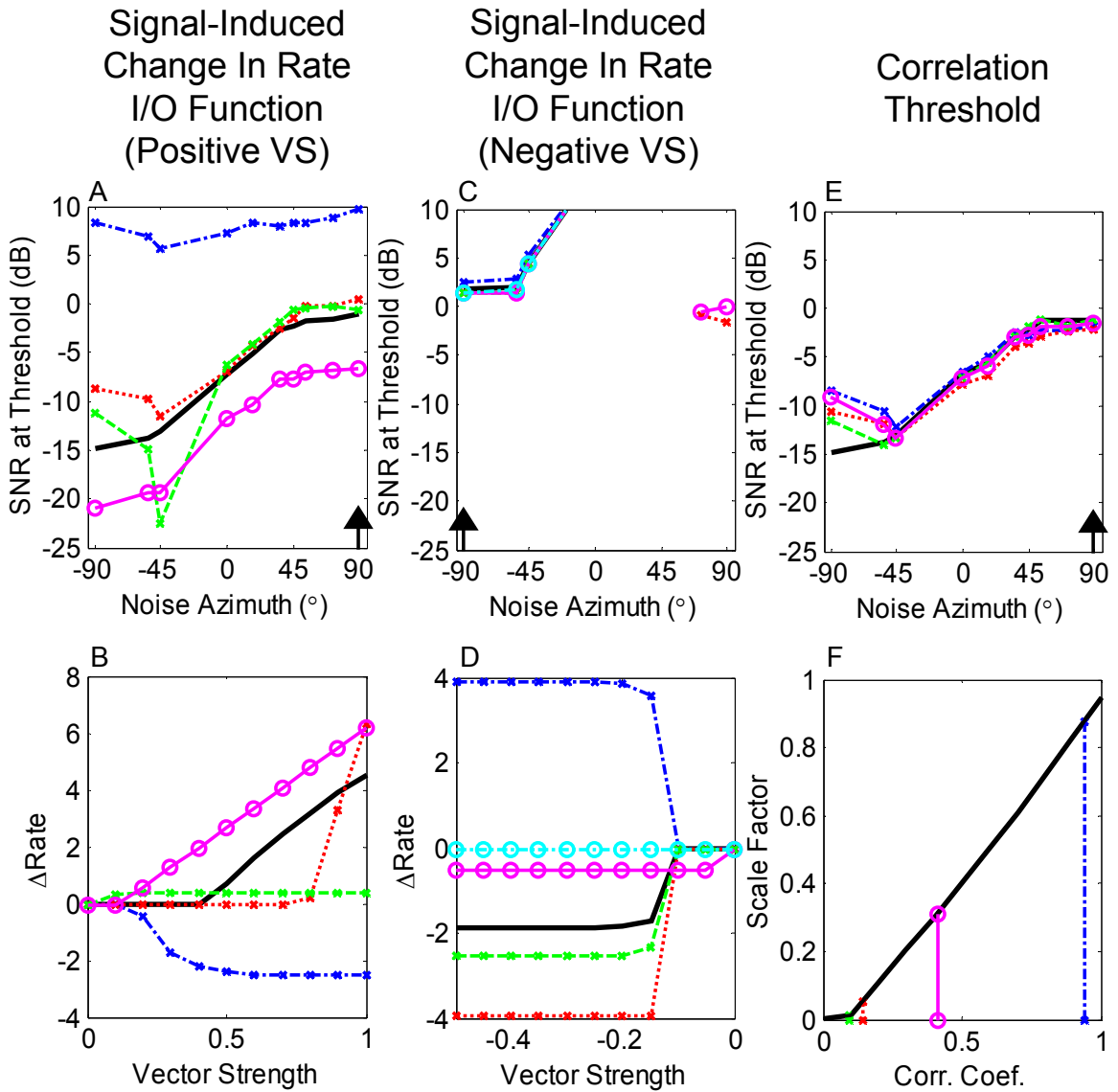


Figure 18: Effect of the signal-induced rate change. (A) Thresholds for a model unit for different signal-induced rate-change input-output functions. All other parameters were set to those of unit 22-0. The signal, denoted by the arrow, is at +90°. The thresholds for the actual input-output function for this unit is shown by the thick black line. (B) The signal-induced change-in-rate input-output functions for positive vector strengths were selected from those fit to the neurons in the population to show the widest variety of thresholds. The thick black line shows the actual input-output function for this unit. (C, D) The same as A and B, except for changes in the input-output function for negative vector strengths and the signal at -90. These functions have little effect on the threshold, although large negative changes or ones that are negative near 0 allow the signal to be detected for the noise at +90° and +90°. (E, F) The effect of the change-in-rate threshold. For correlations below threshold, no change in rate is allowed (see text). For small correlation thresholds, the change in rate is always allowed, and the thresholds are low for negative azimuths. As the correlation threshold increases, the thresholds tend to show a “bounce” for negative azimuths because the correlation is pulled below threshold at these azimuths. The true correlation threshold for this unit was 0, which is difficult to see, but gives no bounce (thick black line in E).

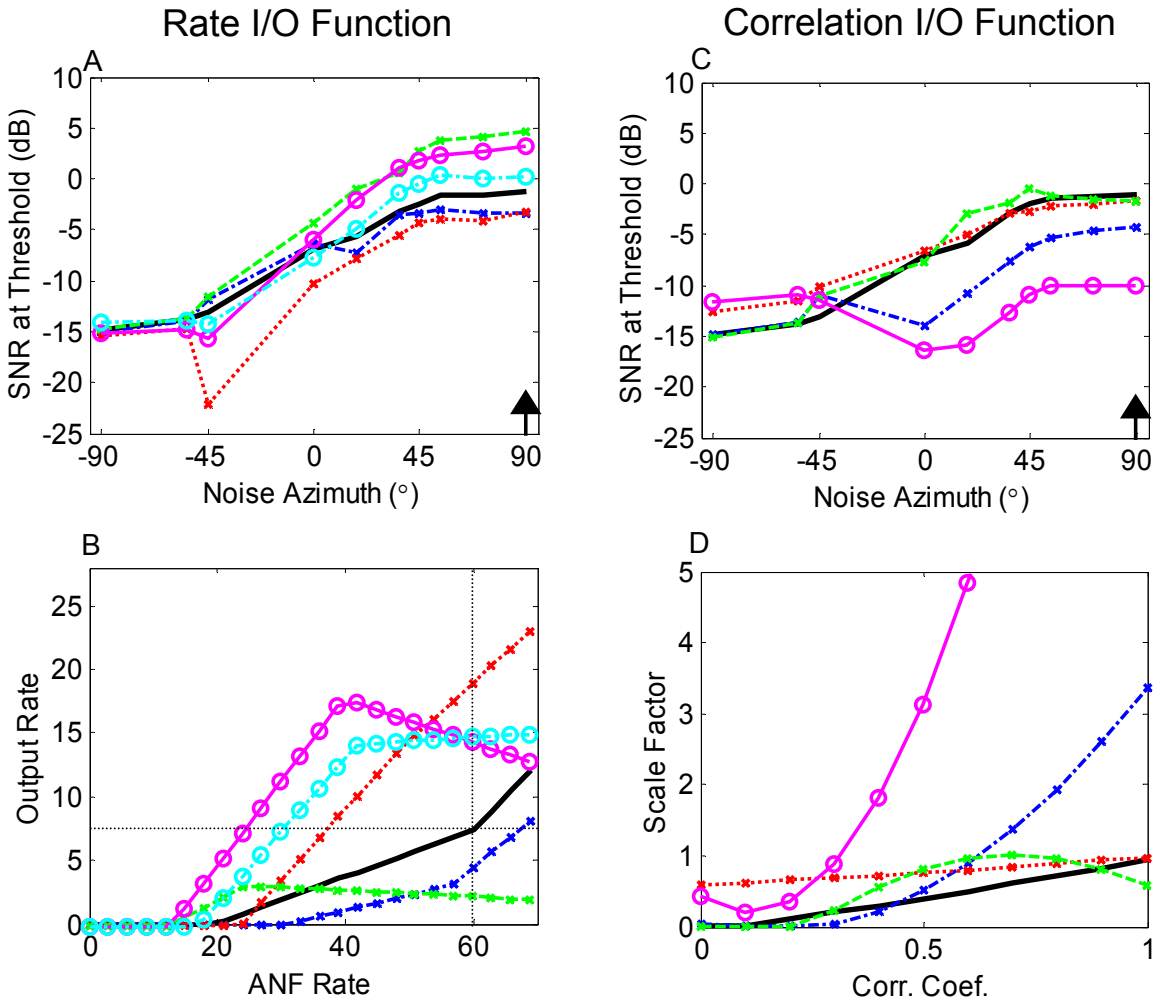


Figure 19: Effect of rate and correlation input-output functions. (A,B) Similar to Figure 18, except for changes in the rate input-output functions. Thresholds are poor when the second line has a zero or negative slope. The best threshold occurs for a rate input-output function with a steep slope for most input rates. Dotted lines in B show the approximate signal-alone input and output rates. (C,D) Same as A and B for correlation input-output functions. The thresholds are best when the slope of the function sharply increases with increased correlation.

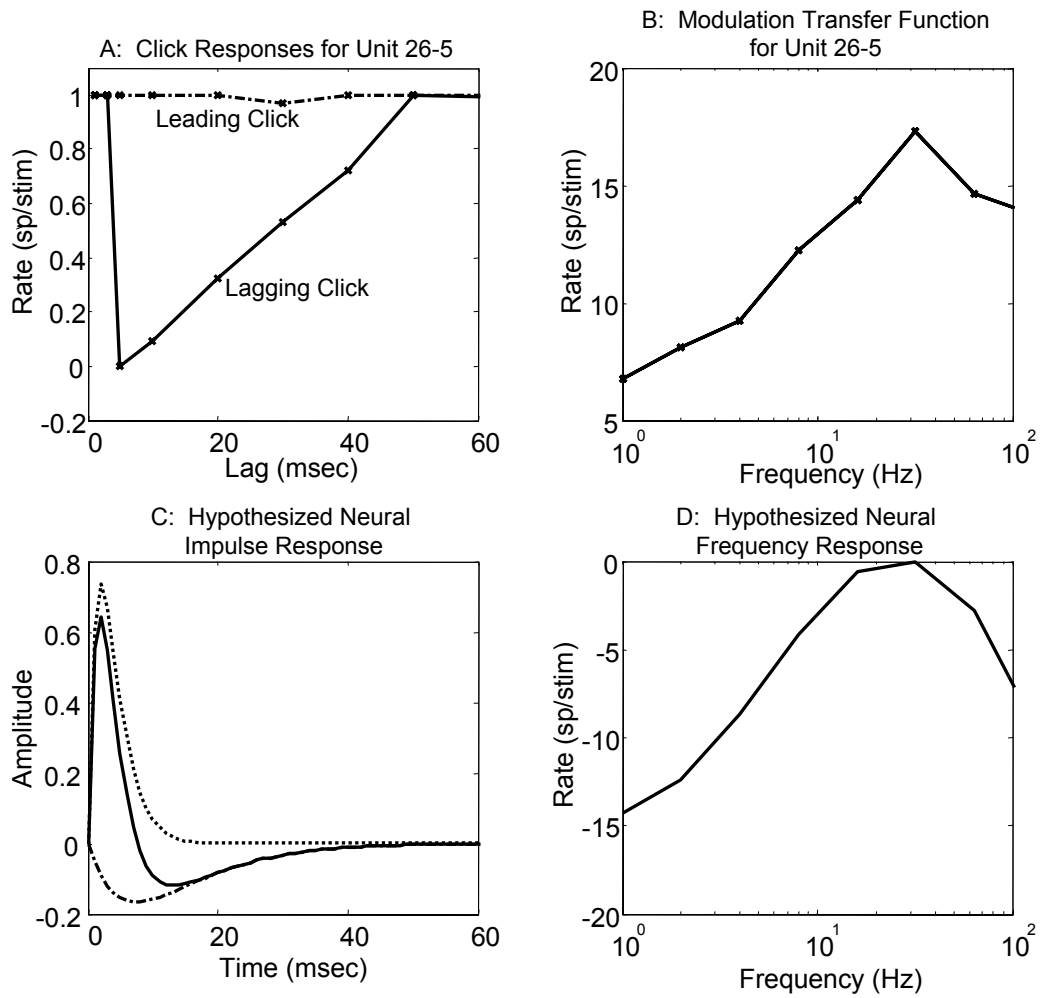


Figure 20: (A) Responses of two clicks where one click lags the other in time as a function of the lag time for unit 26-5. The leading click can suppress the lagging click response for several msec. (B) Modulation transfer function for unit 26-5 recorded using AM modulated noise bursts with different modulation rates. This unit shows an enhanced response to 30 Hz modulation rates. (C) Impulse response of a hypothesized neural filter composed of a large and quick EPSP (dotted lines) and a slow, smaller IPSP (dash-dot lines). (D) Frequency response of the hypothesized neural filter. This hypothesized filter would give similar responses to those seen in Panels A and B.

Chapter 4: Human Behavioral Thresholds Compared to Neural and Model Population Thresholds

Introduction

Previous psychophysical studies have shown that spatial release from masking for low-frequency stimuli is due at least in part to binaural processing (Gilkey, Good, and Ball; 1997), and the binaural cues involved are thought to be interaural time differences (ITDs). In the first chapter, we showed how low-frequency, ITD-sensitive units in the inferior colliculus (IC) respond to signals and maskers at different spatial locations, and in the second chapter, we presented a model of the single unit responses that reproduces the rate responses and thresholds of the single units reasonably well. For both the data and the model, the single-unit thresholds depend on the unit's best and worst azimuths as well as the location of the signal and masker, rather than on the separation of the signal and masker per se. Consequently, in order to find a correlate of spatial release from masking, we must look at the neural population as a whole, which would include units with a variety of best and worst azimuths. If such a correlate exists for both the neural and model populations, then we will have shown that these units are likely to be important for low-frequency spatial release from masking, and we will have a computational model that is able to predict both cat individual neuron and human behavioral data. In this chapter, we first present human psychophysical thresholds measured using stimuli similar to those used in our physiological experiments. We then compare population response from the single units from Chapter 2 and the model units from Chapter 3 to the human behavioral thresholds. In Chapter 3, we presented two models, the modified cross-correlator model, which was similar to the cross-correlator models of Jeffress (1948) and Colburn (1973, 1977a, b), and the envelope-processor model, which added sensitivity to the stimuli's amplitude modulation. In the same way that the modified cross-correlator model did not predict the individual unit responses, we show in this chapter that adding this sensitivity to the stimuli's envelope is necessary to predict the neural population response and the human psychophysical thresholds. Additionally, we show that if the model units' characteristic frequency (CF) and best

interaural time delay (ITD) are independent, then best thresholds for the entire population do not match the psychophysical thresholds. In contrast, if the units' CFs and best ITDs have an inverse relationship similar to the one seen in the IC of guinea pigs (McAlpine et al., 2001) and cats (Joris et al, 2003), the population response better matches the psychophysical thresholds, providing indirect evidence that the inverse relationship between an individual unit's CF and ITD seen in other mammals may also hold in humans.

Methods

Spatial release from masking was measured for three female and two male normal-hearing human subjects using lowpass stimuli. (Thresholds for broadband stimuli were also measured, and the results are reported elsewhere; see Lane et al., 2003.) For these experiments, the signal was a 200-ms 40-Hz chirp train, bandpass filtered between 200 and 1500 Hz, and the Gaussian noise masker was bandpass filtered between 200 and 2000 Hz. Stimulus azimuth was simulated using human non-individualized head-related transfer functions (Brown, 2000), and stimuli were delivered via insert earphones to subjects in a sound-treated booth. The signal was fixed in both azimuth and spectrum level (14 dB re 20 μ Pa/ $\sqrt{\text{Hz}}$). For each masker azimuth, the masker level was adjusted to find the detection threshold using a 3-down, 1-up procedure to estimate the signal-to-noise ratio (SNR) yielding 79.4% correct detection performance (Figure 1). The physiological and modeling methods are the same as those discussed in the previous two chapters.

Results

Human behavioral thresholds and neural population thresholds

Figure 1 shows the human psychophysical thresholds as a function of noise azimuth for three signal azimuths (arrows). These results show spatial release from masking in that the thresholds generally improve as the noise is moved away from the signal. We would like to see if the low-frequency, ITD-sensitive units in Chapter 2 are sufficient for explaining the low-frequency spatial release from masking. As shown in Chapter 2, individual unit thresholds change with signal and masker azimuth; however, the changes

in threshold are related to the signal and masker locations with respect the unit's best and worst azimuth, rather than just the signal and masker's separation. We hypothesize that, for the low-frequency, ITD-sensitive units described in Chapter 2 to provide a neural substrate for low-frequency spatial release from masking, their responses must be combined in a population response. Here, we define the population threshold in a simple way: for each signal and noise configuration, the population threshold is the best single-unit threshold in our sample of ITD-sensitive units. Figure 2 shows the individual unit thresholds (dot-dash lines) and the population thresholds (solid lines, shifted down 2 dB to reveal the single unit thresholds) as a function of noise azimuth for three signal azimuths (arrows). Unlike single unit thresholds, the population thresholds do show spatial release from masking because the population thresholds generally improve when the signal and noise are separated.

Model population thresholds

We also constructed a population threshold based on the model units described in Chapter 3. Specifically, we used the “envelope-processor model,” which consisted of three signal pathways: a rate-level processor which gives rate-level functions with varying shapes, a cross-correlation pathway that gives sensitivity to ITDs, and an envelope processor which provides sensitivity to the input's temporal envelope. The predictions from this model are compared to the “modified cross-correlator model,” which does not include the sensitivity to the modulation. Here, we first look at the population response of the two sets of model units, where the model parameters were fit to each individual unit separately. The fits were unaltered from those shown in Chapter 3, i.e. the model parameters are fit to predict the mean rate responses, not the thresholds themselves. Figure 3 shows the modified cross-correlator population responses in a format similar to that shown in Figure 2. The threshold curves for many of the individual units are sharply non-monotonic, especially for the signal at $+45^\circ$ and $+90^\circ$. As a result, the population thresholds (thick black lines) do not improve with signal and masker separation as smoothly as the neural population and human behavioral thresholds do.

The population thresholds for the envelope-processor model are shown in Figure 4. Although many of the individual unit thresholds are still non-monotonic, the non-

monotonicities are less common and less severe. The population thresholds for the envelope-processor model generally do show spatial release from masking, except for the signal at -90° (see below), and this model gives thresholds that are similar to the neural population. In general, both the individual unit thresholds and the population thresholds for the envelope-processor model are more similar in shape and magnitude to the neural population than the modified cross-correlator model (see also Figure 5). Consequently, consistent with the results from the previous chapter, the units' sensitivity to envelope modulation appears to be critical for predicting low-frequency signal detection for these units and stimuli.

Additionally, for the two model populations, we show the entire population's response for the signal at -90° as well as $+90^\circ$. -90° is an unfavorable azimuth for most of the units in the population (in this case, the signal is in the ipsilateral hemifield). Many of the thresholds at -90° could not be measured for either of the two models because the signal could not be detected at any noise level, which is why there are fewer individual thresholds curves for this signal azimuth. Additionally, the population thresholds do not show SRM for the signal at -90° : when the separation between the signal and the masker increases beyond 90° , the population thresholds get worse rather than better.

For each of these two models, we show the population thresholds for the signal at $+90^\circ$ (red dotted line) on the same figure as the one with the signal at -90° , but reflected about the midline. Combining these two populations is equivalent to placing the signal to the animal's right and comparing the units in the left IC (population threshold for S $+90^\circ$) to the units in the right IC (population threshold for S -90°). For both models, the population thresholds for the left IC (contralateral to the signal) are better or about the same as those in the right (ipsilateral) side, indicating that the units for which the signal produces an excitatory response have better thresholds for these stimuli, consistent with results in the previous chapters. Furthermore, consistent with the results shown in Chapter 3, Figure 15, the S -90° , N $+90^\circ$ (i.e., the signal at -90° and the noise at $+90^\circ$) thresholds for the modified cross-correlation model are often better than those for the S $+90^\circ$, N -90° thresholds for individual units; however, the modified cross-correlator model population thresholds are about the same for the two conditions. In contrast,

adding the envelope processing generally makes the S+90°, N-90° thresholds better for the individual thresholds; the population thresholds for the envelope-processor model are better in the S+90°, N-90° condition compared to the reverse by more than 10 dB (red dotted curve compared to black solid curve).

Figure 5 compares the low-pass human psychophysical thresholds (A) to the cat neural population thresholds (B) and the two model population thresholds (C, modified cross-correlator model, and D, envelope processor model). We have assumed that the psychophysical thresholds will be symmetric for the case when the signal is at 0° and have reflected the thresholds about the y-axis (thin dash-dot line) to aid visual comparisons. In order to compare the human behavioral thresholds to the cat neural population thresholds despite the difference in species head size, the axes are matched for noise ITD (lower axis) rather than noise azimuth (upper axis). The psychophysical thresholds and the neural population thresholds have similar shapes, amounts of spatial release from masking, and actual threshold values, although the neural population thresholds appear to fall off more steeply with separation of signal and masker and have higher thresholds when the signal and masker are co-located. Accordingly, these similarities suggest that low-frequency, ITD-sensitive units in the IC may be important for the binaural component of low-frequency spatial release from masking.

The population thresholds for the envelope-processor model (Panel D) match the neural population thresholds (Panel B) better than the modified cross-correlator thresholds (Panel C) in that the shapes and magnitudes of the thresholds of the curves are better matched. Also, the population thresholds for the modified cross-correlator model (Panel C) for the noise at -290 μ sec are all about the same, regardless of the signal azimuth, which is not consistent with the neural population results or with the idea that the thresholds improve as the signal and masker are separated in space. Because the envelope-processor model population thresholds are similar to those of the neural population, the differences seen between the neural population and the human behavioral thresholds are also seen for the envelope-processor model population. However, to the extent that the model population response predicts the human behavioral thresholds, it appears that low-frequency spatial release from masking in humans depends on both interaural cross-correlation *and* sensitivity to envelope modulation.

CF and best ITD joint distribution affects population thresholds

In Chapter 3, we showed that the CF and the best ITD of a unit largely determine the shape of its threshold curve because these parameters determine the best and worst azimuths, and thereby determine the best and worst thresholds. In Figure 6, we explore a population of model units where all of the model parameters are fixed except for their CF and best ITD. In the previous figures, we matched the model parameters for each model neuron to those of a neuron in our IC population; consequently, the best ITDs and CFs for the units in Figures 3-5 have an inverse relationship as noted by McAlpine et al. (2001) in guinea pigs and as shown in Chapter 2. It is suspected that other mammals, including humans, may have such an inverse relationship, but it is difficult to directly determine this relationship in humans. The relationship between CF and best ITD is especially interesting because the Jeffress (1948) and Colburn (1973, 1977a, b) cross-correlator models assume a set of units that compose a full matrix of every combination of best ITD and CF. Interestingly, our analysis of the psychophysical results gives some insight into the CF-best ITD relationship in humans.

Figure 6 shows the effect of having neural populations with two different joint CF-best ITD distributions using the envelope-processor model. We have fixed all of the model parameters those for one unit (unit 22-0 from the previous chapters) except the unit's best ITD and CF. This unit's parameters give some of the best thresholds in the population, although not when the signal and masker are co-located. Panel G shows the combinations of best ITDs and CFs used (colored dots). In Panels A and D, we show the human behavioral thresholds from Figure 5 for the signal straight ahead (Panel A) and to the right (Panel D). In Panels C and F, we show the model unit thresholds and the population thresholds for all the units in Panel G. Unlike the model population thresholds shown in Figure 5, the shapes of the model population thresholds are not that similar to the behavioral thresholds. For the signal at 0° (Panel C), the population thresholds fall off steeply with separation of signal and masker. We show the population thresholds reflected about the midline (dotted line) to account for the presence of both the left and right ICs, but the thresholds are even steeper. In Panel B, we show the population threshold when only the units that were inside the dashed black lines in Panel G were included in the population (Panels B and E). The black lines represent the

constant best IPD lines for 0.1 and 0.3 cycles, which contain most of the best ITDs and CFs of the actual units in our sample (black x's). In this case, excluding the units with potentially non-physiological CFs and best ITDs degrades the predicted performance when the masker is in the contralateral hemifield. However, when the population threshold was reflected about the midline, simulating the effect of having two ICs, the population thresholds become similar to that for the full set of units, indicating that, once units in both ICs are included, no conclusions can be made about the availability and usefulness of neurons with different CF and best ITD combinations for the signal straight ahead. As shown in Chapter 3, the thresholds when the signal and masker are co-located can be improved with different parameter settings, and because these predictions are generally too high when the signal and masker are co-located, it is possible that this unit's parameters may not have been the optimal choice for this simulation.

For the signal to the right (Panel F), the thresholds for the full population again do not show the correct shape. In this case, the thresholds do not improve monotonically as the noise masker is separated from the signal. Instead the best thresholds occur for the masker straight ahead (at 0 μ sec). The units that cause the population thresholds to be better for the masker at 0 μ sec are those with high CFs and large best ITDs. Panel E shows the effect of excluding units with large and small best IPDs (those outside the dashed lines in Panel G); in this case, the thresholds are a reasonable match to the behavioral thresholds, except that the co-located thresholds are somewhat too high, as in the actual neural population. Including units with best IPDs up to 0.4 cycles (rather than 0.3 cycles as in Panel E) gave poor predictions, similar to those in Panel F. In order to predict the psychophysical behavior of the human listeners, we had to exclude units with best IPDs greater than 0.3 cycles, perhaps indicating that human listeners either do not have or cannot use neurons with large best IPDs.

The reason that units with large best ITDs and high CFs give such low thresholds is that they have their best and worst azimuths close together. For these high CF units, thresholds change rapidly with noise azimuth, making the best thresholds (at the worst azimuths) fall fairly close to the signal azimuth (at the best azimuths). In contrast, for the inverse-relationship array, units with best azimuths near the signal at +90° have slowly changing thresholds due to their low CFs. Consequently, their best thresholds (and worst

azimuths) lie somewhere in the ipsilateral hemifield. For this type of cross-correlator array to predict the slowly changing thresholds seen in the psychophysical experiments, there cannot be units with large best ITDs and high CFs affecting the population response.

Figure 7 makes a similar point, using a more abstract model that 1) can be applied more directly to the human psychophysical data, 2) requires much less computational power, and 3) depends only on the best ITD and CF of the units, removing the effects of the other parameters of the full neural model. As shown in Chapter 3, Figure 16, the shape of the threshold curve for each unit is predicted fairly well by a cosine that has a peak at the best ITD and a valley at the worst ITD. This abstraction is likely to hold for most neural models with cross-correlation because the thresholds are likely to have similar shapes regardless of the implementation details. Two possible arrays of units with different CF and ITD relationships are again shown. For the independent array (right panels), only three best ITDs were used: 725, which corresponds to $+90^\circ$ in our human HRTFs (black x), 825, which is just outside the physiological range (red circles) and 625 μ secs, which corresponds to about 77° (blue diamonds); units had CFs ranging from 100-1600 Hz. These units represent a subset of the entire independent array of units with every combination of CF and ITD, but since these units have best azimuths near $+90^\circ$, these are the ones most likely to give the best thresholds when the signal is at $+90^\circ$ (see Chapters 2 and 3). For the inverse-relationship array (middle panels), the same CFs were used, but the three rows were determined by constant best IPDs of 0.2 (black x's), 0.3 (red circles), and 0.45 (blue diamonds). In contrast to the independent array, these units would represent a large fraction of the units seen in the population if there were indeed an inverse relationship between the CF and best ITD. The black dotted line shows the physiological range as defined by $+90^\circ$ in our human HRTFs. Only the units have best ITDs at or outside the physiological range were included in this analysis because these are the units that have $+90^\circ$ as their best azimuth, making these units the ones that produce the best thresholds.

Figure 7 shows these cosines for the units with the given best ITDs and CFs. The worst threshold for each cosine, assumed to occur for the signal and noise co-located at $+90^\circ$, was normalized to 1, and the best threshold (the cosine's minimum) is normalized

to 0, although this minimum can (and often does) fall outside of the physiological range. This normalization procedure was based on the results shown in Chapter 3, Figure 16, where the co-located thresholds for all the various combinations of best ITD and CF were all similar. To facilitate visual comparisons, we only plot thresholds at noise azimuths for which we have behavioral thresholds. For the independent array in the right panels of Figure 7, the best thresholds can occur at small separations of signal and masker because all CFs are allowed for the best ITDs used. Thus, for even some of the smallest separations of signal and noise (less than 45°), the population thresholds (thick black line) are as good as any of the others.

For the inverse-relationship array (middle panels), the best thresholds more closely match those seen in the psychophysical results because they change more slowly with noise azimuth. Interestingly, the thresholds corresponding to a constant best IPD of 0.2 cycles (black x's), which are the most frequent in the cat results of Hancock and Delgutte (2003), do not give results that match the human psychophysical data. Instead, units with a best IPD of 0.45 cycles seem to give the best predictions of the psychophysical data, although the 0.3 units provide good thresholds for separations greater than 90° . Allowing best IPDs of up to 0.5 cycles still gives reasonable match to the psychophysical data, but 0.6 cycles gives thresholds that improve too quickly with separation (not shown). This value for the highest allowed best IPD is different from the highest allowed best IPD found in Figure 6. There are many differences between the two figures: (1) we are using a different set of human behavioral thresholds (the signal placed at $+90^\circ$ as opposed to $+30^\circ$), (2) we are using two different models of thresholds, and (3) we are not sampling the CFs in Figure 6 as closely as we did in Figure 7. Consequently, we are unable to give a confident estimate of the highest best IPD used by humans based on this set of psychophysical thresholds; however, this analysis does suggest that such an upper limit does exist, a result which has not been previously shown.

Discussion

We have shown that low-frequency, ITD-sensitive units investigated in this thesis can provide a neural substrate for low-frequency spatial release from masking. The population responses for both the individual units and the model units show spatial

release from masking that resembles that seen in human psychophysics. In addition to the well-known cross-correlation mechanism, our results suggest that successful prediction of the results depends on a mechanism that increases IC responses to the signal over the noise, presumably due to sensitivity to envelope modulation. Therefore, low-frequency signal detection in humans appears to rely on both interaural cross-correlation and sensitivity to envelope modulation for some stimuli.

Additionally, the modeling results suggest that the CF and best ITD are inversely related for the units making up the population. Without this relationship, which is similar to the results seen physiologically in guinea pigs and cats, thresholds for small signal and masker separations are better than those obtained for human behavior.

As shown in the previous chapters, the best thresholds occur when the signal is at a favorable azimuth and the noise is at an unfavorable one. This result is different from the results suggested by the differences in thresholds seen when comparing the $N0S\pi$ to the $N\pi S0$ thresholds, where the most sensitive units seem to be the ones where the signal is placed at an unfavorable IPD. Unlike the previous chapters, here we showed that this discrepancy holds with or without envelope processing for the population response. It seems that the signal is detected by more units when placed at a favorable azimuth. However, the difference in the population thresholds for the $S+90^\circ$, $N-90^\circ$ thresholds compared to $S-90^\circ$, $N+90^\circ$ is much larger for the envelope processor model than for the modified cross-correlator model (Figures 3 and 4). Furthermore, the population thresholds for the signal at -90° do not show a correlate of spatial release from masking, and consequently the units ipsilateral to the signal location are probably not the units used to detect the signal. This hypothesis could be tested in patients or animals with a lesioned IC. For these stimuli, we would predict that a subject with a lesioned right IC would show normal spatial release from masking when the target was straight ahead or to the right of the subject, but not to the left. Interestingly, Litovsky et al. (2002) measured spatial release from masking in a patient with a unilateral IC lesion, and the patient showed no impairment. However, in this study, the subject was only tested when the signal was straight ahead so that these results do not test our hypothesis directly.

In summary, we have shown that both a neural population and a model population can predict spatial release from masking similar to results seen psychophysically. The

model responses require ITD processing via interaural cross-correlation, as expected, but must also include envelope processing, indicating that both of these mechanisms may be involved in signal detection by the human auditory system.

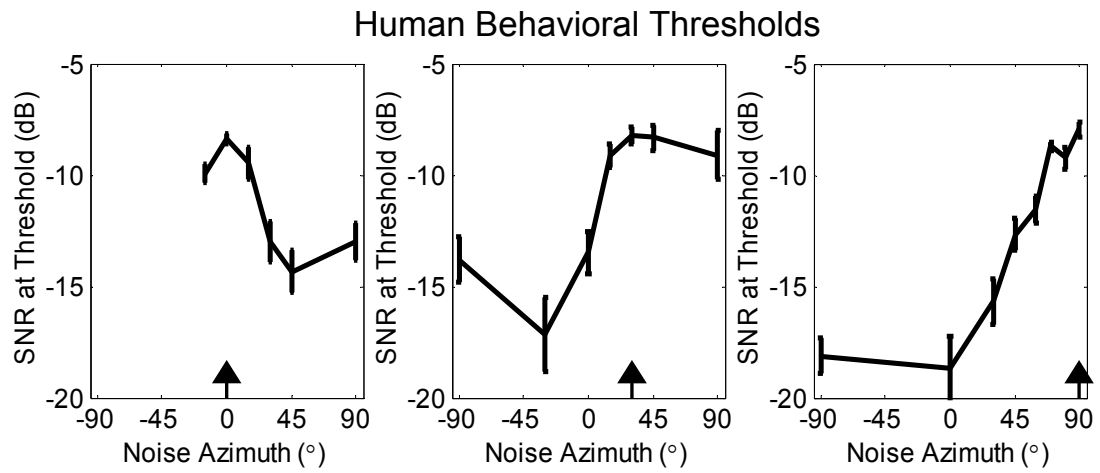


Figure 1: Human behavioral thresholds averaged across five subjects for low-pass chirp signals for three signal azimuths (arrow). Error bars show ± 1 standard error across subjects.

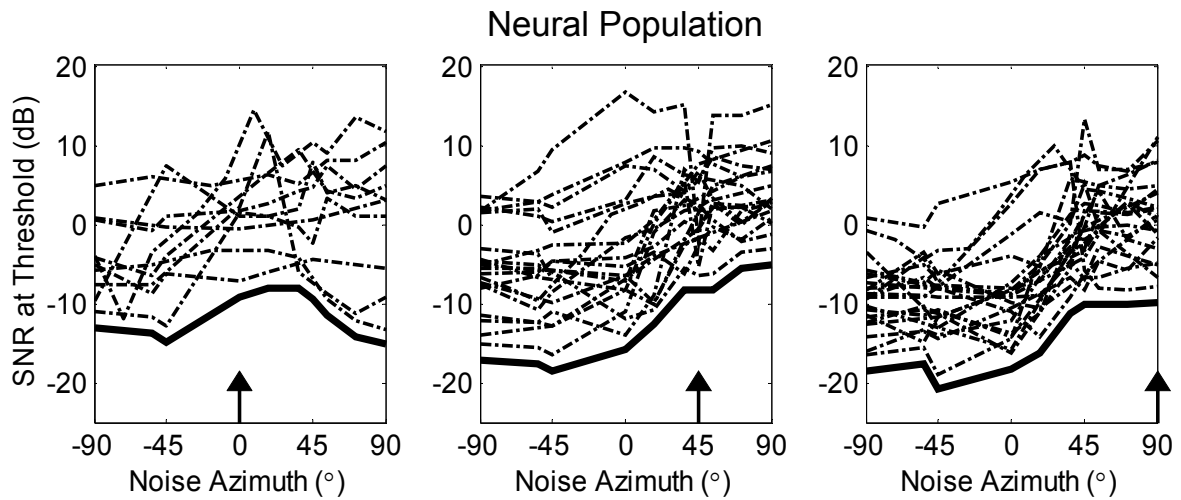


Figure 2: Neural population thresholds for three signal azimuths (arrows). Dash-dot lines: single unit thresholds; solid lines: population thresholds (offset by 2 dB).

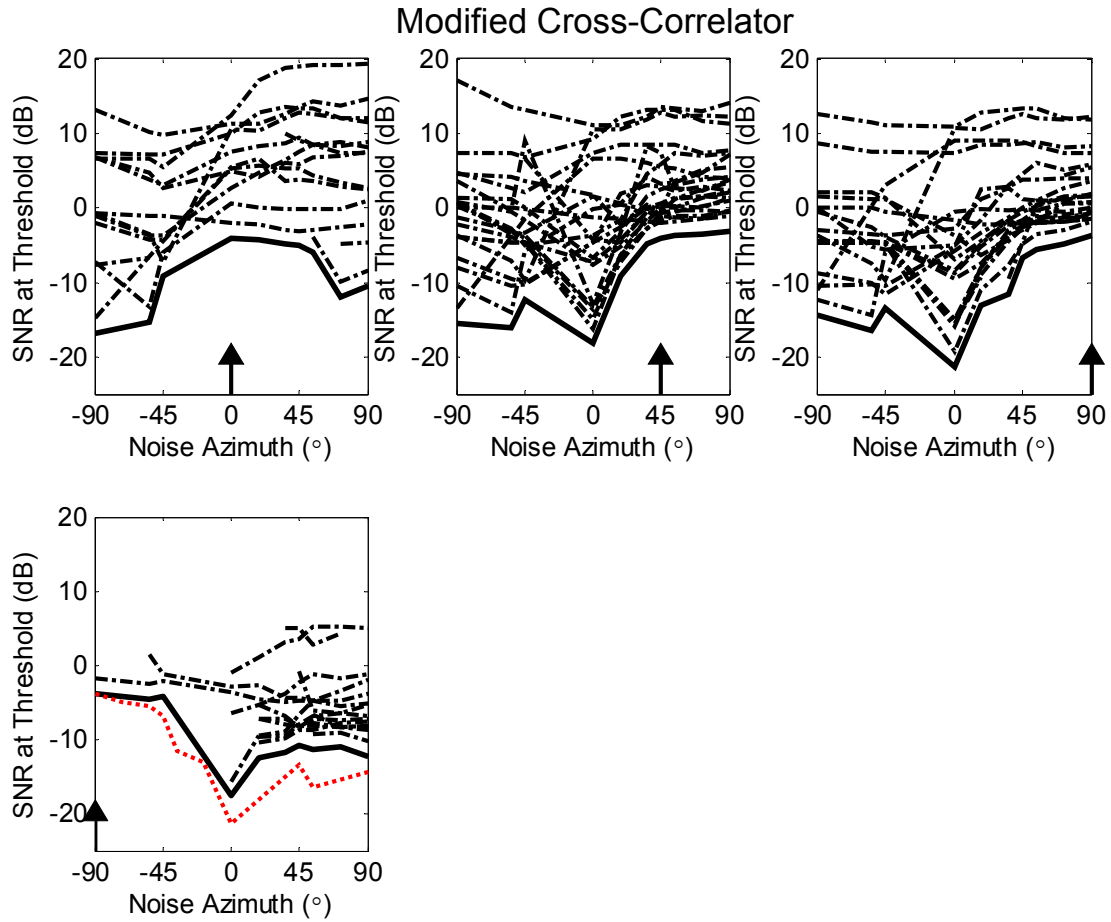


Figure 3: Population thresholds for modified cross-correlator model for four signal azimuths (arrows). Same format as Figure 2. Many of the individual units show sharply non-monotonic threshold functions, which results in population thresholds that do not always improve with separation of the signal and masker. For the signal at -90° , for many of the model neurons, the signal could often not be detected for any noise level, resulting in fewer plotted thresholds. Red dotted line shows the threshold curve for the signal at $+90^\circ$ reversed in azimuth, comparing the thresholds for the units in the left IC and the right IC. The thresholds for the units in the IC contralateral to the signal are either about the same or better than the units in the ipsilateral IC.

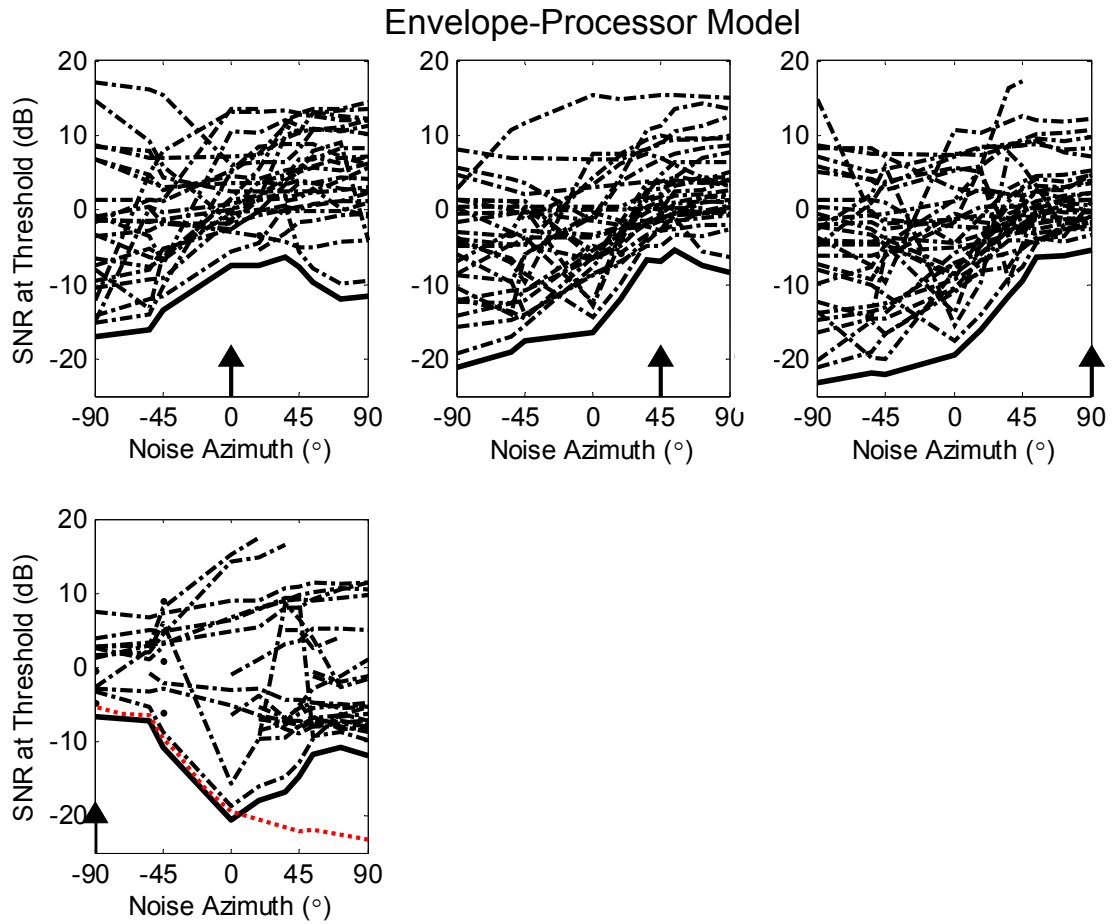


Figure 4: Population thresholds for the envelope-processor model for four signal azimuths (arrows). Same format as Figure 3. More of the units show monotonic threshold curves, resulting in thresholds that improve as the signal and masker are separated, except for the signal at -90° . Again, the thresholds for the units in the IC contralateral to the signal are either about the same or better as the units in the ipsilateral IC.

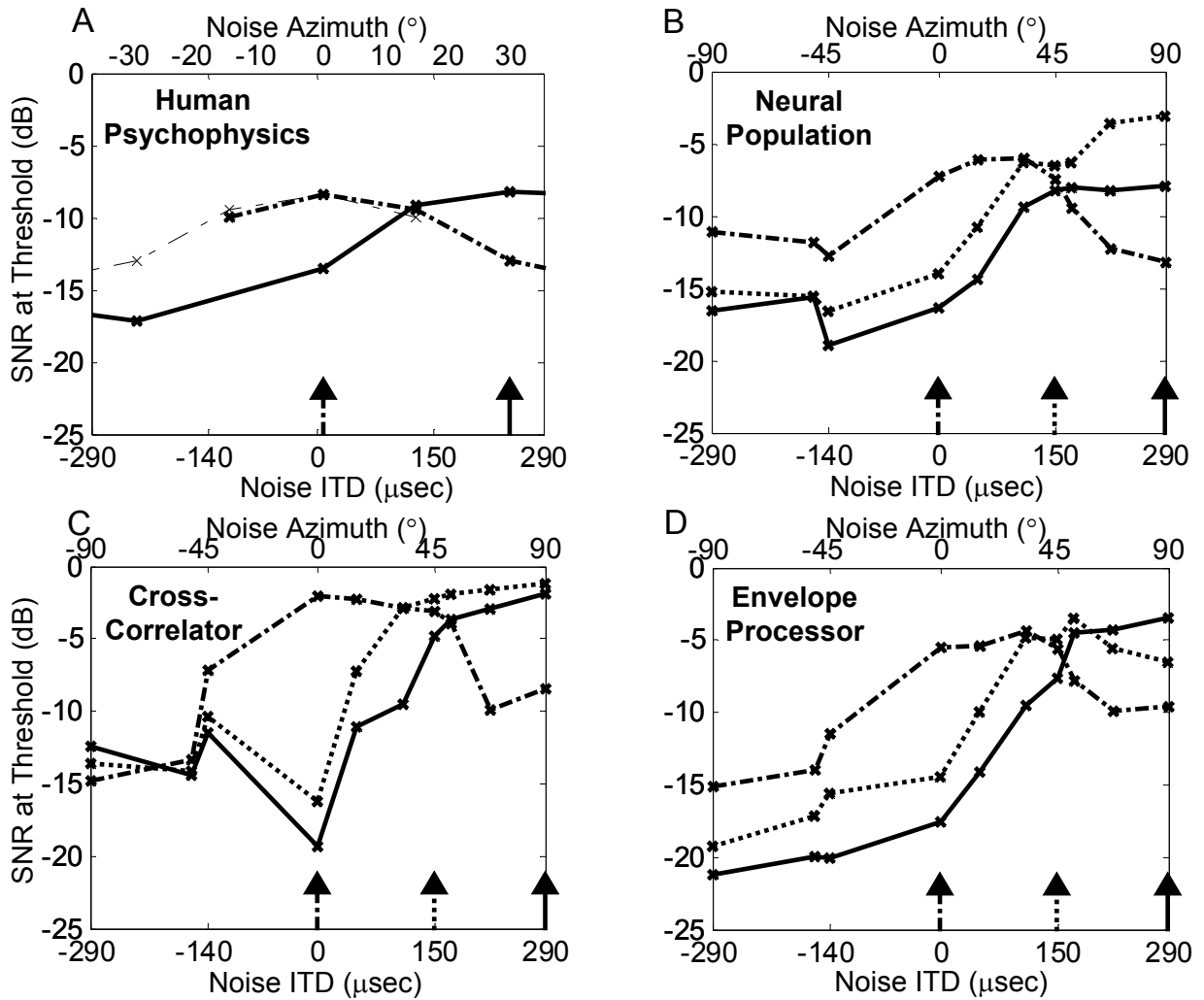


Figure 5: A) Human psychophysical thresholds for low-frequency stimuli as a function of noise ITD (lower axis) and azimuth (upper axis). Arrows indicate signal azimuth, arrow tail indicates corresponding threshold curve. B) Neural population thresholds C) Population thresholds for the modified cross-correlator-model population thresholds D) Population thresholds for the envelope-processor-model

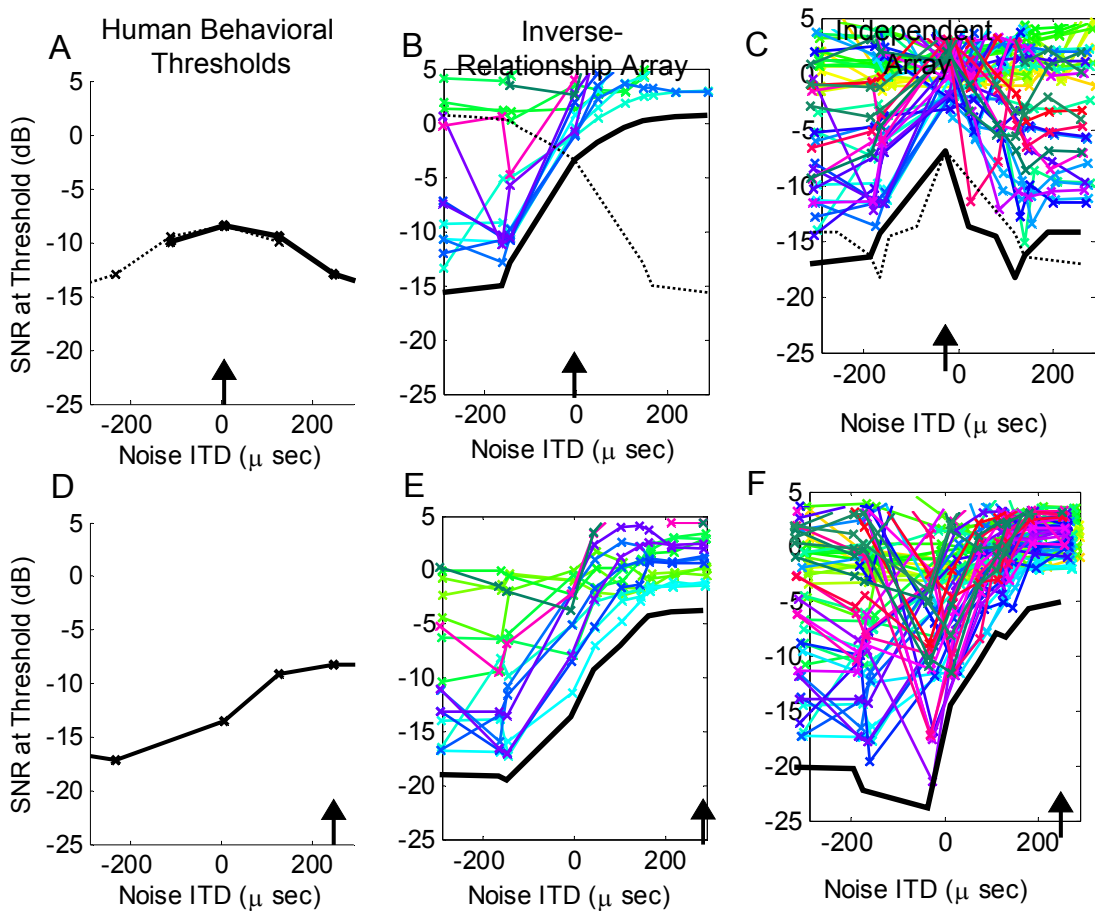


Figure 6: Model populations for two different distributions of CF and best ITDs. A) Human behavioral thresholds for the signal at 0°. B and C) Individual unit thresholds (color) and population thresholds (thick solid line) for the signal at 0°. The dotted black line is the population threshold curve reflected about the y-axis, representing the model responses for the units in the other IC. The model parameters were fixed for unit 22-0 except the best ITD and CF. In C the full array of best ITDs and CFs were used (see G), while for B, only model units with CFs and best ITDs that fell within the dotted lines in G were used. D-F) Same as A-C for the signal at +90°. G) The array of the units' best ITDs and CFs. The black x's represent the data points in our actual population.

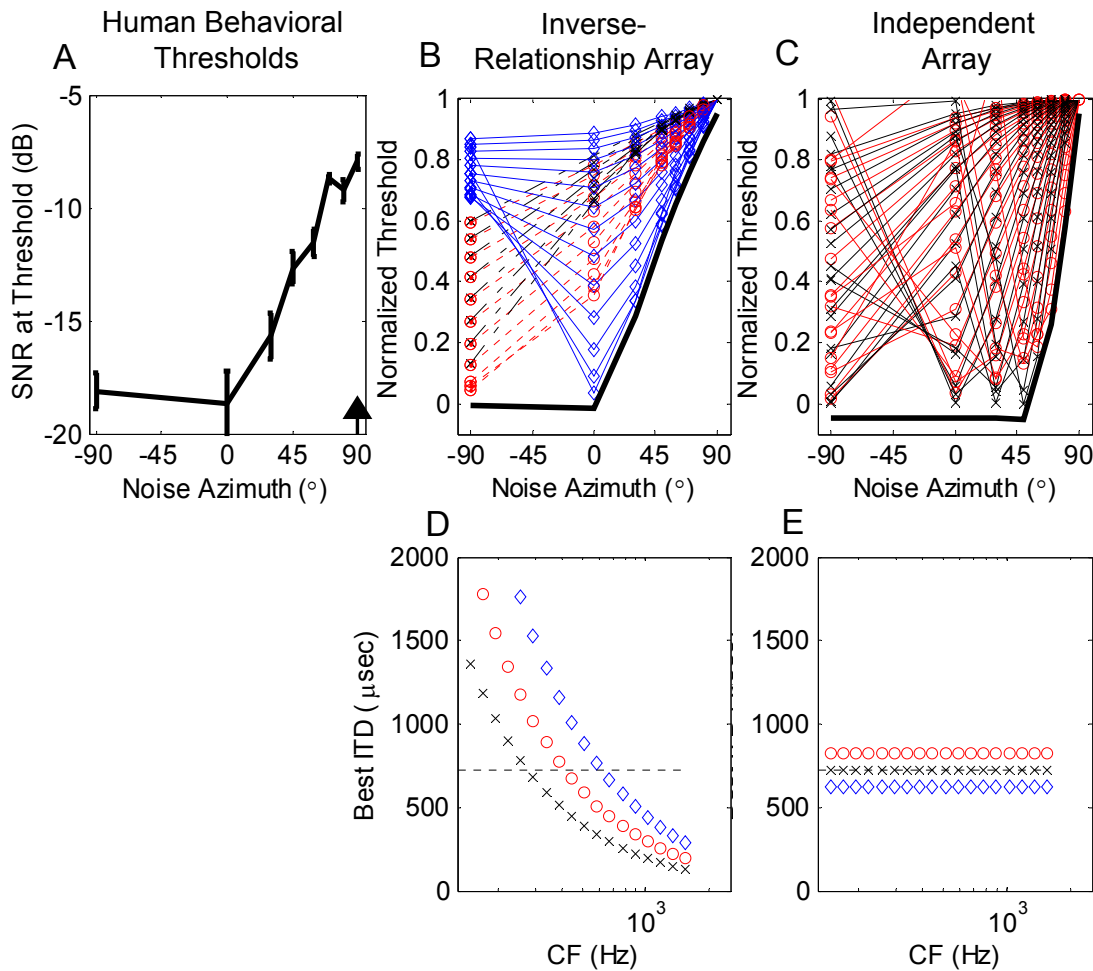


Figure 7: Hypothetical population responses for two different unit populations. Signal is placed at +90°. In this figure, all azimuths are defined by the human HRTFs. A) Human behavioral thresholds. B) Normalized thresholds for the units in D above the black dotted line. Thick black line is the population threshold (offset by 0.05). No data points were plotted between -90° and 0° to match the available psychophysical thresholds. C) Normalized thresholds for the units in E, otherwise same as B. D) Units with a constant best IPD of 0.2 (black x's), 0.3 (red circles), and 0.45 (blue diamonds). E) Units with three different best ITDs near +90°: 625 μsec (blue diamonds), 725 μsec (black x's), and 825 μsec (red circles).

Chapter 5: Conclusions

In this final chapter, we first summarize the results from the previous chapters and discuss some of the consequences of the methods used in this thesis. We then discuss possible neural implementations for the processing paths included in our neuron model and propose specific physiological experiments that could improve our understanding of these paths' actual neural implementation. We then discuss how a central processor could detect signals in noise based on the processed information from the IC. We also give a brief discussion of future psychophysical experiments that could test the interaction of binaural and envelope processing in human listeners, and we finish by describing how the results of this thesis might be used to improve hearing aids and automatic speech recognition systems.

Thesis Summary

In Chapter 2, we described how low-frequency, ITD-sensitive units in the IC respond to a 40-Hz chirp-train signal and a broadband noise masker placed at different spatial locations. We showed that individual units do not show true correlates of spatial release from masking because their thresholds do not necessarily depend on the separation of the signal and masker. Instead, their masked thresholds seemed to depend on the relationship between the signal and masker locations and the unit's best and worst azimuth. The best thresholds for single units for these stimuli occurred when the signal is placed at the best azimuth and the noise is placed near the unit's worst azimuth. Additionally, we compared the single unit responses to predictions from a cross-correlator model: although some units' responses could be predicted by the cross-correlator model, other unit responses showed clear differences, indicating that additional processing appears to influence the responses of these low-frequency, ITD-sensitive units.

We then developed a neuron model (Chapter 3) that describes the rate responses and thresholds for the low-frequency, ITD-sensitive units in the IC. We began with a cross-correlator similar to the one described by Colburn (1973, 1977a, b), but as expected

from the single-unit data, additional processing was required to correctly predict the neural responses. The final model, termed the “envelope-processor model” included three processing paths: a cross-correlator pathway that provided ITD sensitivity, a rate-processing path that gave appropriate rate-level functions, and an envelope-processing path that created sensitivity to the stimulus envelope modulation. The model again showed that the best thresholds occur when the signal is placed at the best azimuth and the noise is placed near the worst. Consequently, a unit’s CF and best ITD, which determine the best and worst azimuths, largely determine the shape of the unit’s threshold curve. The other model parameters, which change the shape of the three memoryless input-output functions, also changed the thresholds: the best thresholds occurred when the unit’s rate increased for even small 40-Hz modulations and when the rate and correlation input-output functions had an upward curvature.

In the fourth chapter, we presented human behavioral thresholds for low-frequency stimuli, which were matched to the stimuli used in the physiological experiments. We computed a population threshold from the individual unit responses, the modified cross-correlator model units, and the envelope-processor model units, using the most sensitive units in the population to determine threshold. We found that the neural population thresholds were similar to the human behavioral thresholds in both magnitude and dependence on signal and masker separation, indicating that low-frequency, ITD-sensitive units in the IC are likely to be involved in low-frequency signal detection. Both model population responses were also similar to the human behavioral thresholds, although the modified cross-correlator thresholds did not predict the neural population thresholds or the human behavioral thresholds as well as the envelope-processor model. These results indicate that both cross-correlation and modulation sensitivity appear to be important for human spatial release from masking. Finally, we showed that the shape of the psychophysical threshold curves seemed to arise from the individual units’ relationship between best ITD and CF, giving indirect evidence for a relationship between these unit characteristics similar to the ones recently reported in other mammals (McAlpine et al., 2001; Joris et al., 2003; Hancock and Delgutte, 2003).

Validity of Methods

In this section, we briefly discuss some of the limitations of our methods and their impact on the results presented in the thesis.

The effects of anesthesia

In the physiological experiments described here, we used barbiturate anesthetized cats. The use of anesthesia is known to change the responses of units in the inferior colliculus, and the effects are generally consistent with increased inhibition, including suppressing spontaneous activity and reducing discharge rates (e.g. Kuwada et al., 1989). Here we examine the potential effects of anesthesia and inhibition in the three pathways in the envelope-processor model.

First, inhibition is known to shape rate-level functions, one of the pathways in the envelope-processor model. Yang et al. (1992) blocked GABAergic inhibition with bicuculline in the anesthetized mustache-bat inferior colliculus. This blocking caused some of the units' strongly non-monotonic rate-level functions to become monotonic or weakly non-monotonic. It seems plausible, then, that there may be fewer non-monotonic rate-level functions in the unanesthetized IC. As we saw in Chapter 3, units with non-monotonic rate-level functions generally have poor thresholds; instead, the best thresholds occur for units with expansive rate-level functions. It is conceivable, then, that the rate-level functions of these more sensitive units may be less affected by the increased inhibition, making the responses of these units, and consequently the population thresholds, relatively less affected by the anesthesia.

The ITD-sensitivity of individual units in the IC is also known to be influenced by anesthesia (Kuwada et al., 1989). The overall rate in the rate-ITD function is lowered, and in some cases, the best ITD and the width of the rate-ITD function can be changed. However, the rate-ITD functions in the unanesthetized rabbit show the same types of damped cosine functions that are characteristic of units with ITD-sensitivity in the anesthetized cat. Therefore, the basic cross-correlator mechanism seems to be a valid descriptor of the responses of ITD-sensitive neurons across species with and without

anesthesia. Because the best and worst ITDs had such a large effect on the individual and population thresholds, knowing how the distributions of these unit characteristics changed in the awake animal would be instructive. It is unclear how this distribution would change, however, because the handful of studies that have described these relationships all used anesthetized animals (McAlpine et al., 2001; Joris et al, 2003; and Hancock and Delgutte, 2003).

The final signal pathway in the envelope-processor model was the envelope processor itself, which changed the overall rate due to the presence of a 40-Hz component in the neural response. We hypothesized that the change in rate due to the stimulus modulation could be obtained through a temporal differentiator, perhaps through timed excitation followed by inhibition. Such a neural filter would give a bandpass modulation transfer function (MTF) like those seen in the IC (Langner and Schreiner, 1988; Krishna and Semple, 2000). In this case, the decreased inhibition due to removing the anesthesia might change the highpass characteristics of the MTF; in the extreme case where the inhibition was completely removed, the bandpass modulation filter would become low-pass, making little difference between the noise response and the 40-Hz chirp-train response. Consistent with these ideas, Caspary et al. (2002) showed that blocking GABAergic inhibition could change the shape of the MTFs in the IC, and in particular, blocking inhibition could make bandpass MTFs become lowpass. This result suggests that the MTFs are at least refined by inhibition at the level of the IC, and this inhibition is likely to be affected by anesthesia. On the other hand, results from multi-unit recordings in the awake squirrel monkey show bandpass MTFs in the inferior colliculus similar to those of studies in anesthetized animals (Müller-Preuss et al., 1994). Furthermore, recent results by Sterbing et al. (2003) show that low-frequency, ITD-sensitive units in the awake rabbit IC show dramatic changes in their responses to binaural beat stimuli when the envelope of the stimuli are amplitude modulated at around 25 Hz. The units can show a large increase or decrease in rate with envelope modulation, similar to the signal-induced rate change that we showed in Chapter 3. Consequently, although the increased inhibition may change the modulation filter characteristics

somewhat, low-frequency, ITD-sensitive units in the IC are sensitive to modulation regardless of the presence of anesthesia. Overall, while the absence of anesthesia would probably change the responses of individual IC units, the mechanisms that we have included in our model all appear to be present and active in both awake and anesthetized animals.

Model implementation

The modeling approach taken in this thesis was dictated by the fact that there were many potential influences involved in shaping the neural responses. Rather than attempting to model all of these influences, we used the model to determine what the most important influences were. Because of this decision, several well-established neural processes were ignored, and the following section discusses some of the effects that these processes could have on the model response.

First, we did not create a model that predicted the responses as a function of time. A model that gave the time-varying rate would allow more interesting insights into the neural responses, and it seems likely that all of the pathways that we included in the envelope-processor model could give complex temporal patterns. However, modeling these complex patterns would require a far more complex model, one that we would have difficulty constraining given the data that we had.

Furthermore, we assumed that the ITD-sensitivity was due to a simple cross-correlation of two auditory nerve fiber responses. It is unlikely that the system is really quite so simple. First, several auditory nerve fibers are thought to converge in the cochlear nucleus (CN) bushy cells on both sides, which in turn projects in the medial superior olive (MSO). This convergence in the CN was ignored in our model as well as the convergence of multiple CN bushy cells onto MSO cells. It is possible that, for example, such convergence could contribute to the modulation sensitivity of the cells. Furthermore, Brand et al. (2002) showed that the cross-correlation at the level of the MSO may be influenced by inhibition, an effect that again was ignored in our modeling efforts, but may prove important for predicting the details of the neural responses.

Probably the least physiological aspect of envelope-processor model implementation was the envelope processor itself. Because we did not know that actual shape of each neuron's MTF, we felt that the change based on the 40-Hz component was our best choice. An actual MTF would give different changes in the unit's response for different modulation rates, rather than relying on 40-Hz alone. We were not able from our data to determine the location or nature of the MTF; however, as discussed below, it seems likely that the MTF is not one filter as implemented here, but a cascade of filters that starts before the cross-correlators in the MSO and is then refined throughout the pathway to the IC.

Definition of single unit and population thresholds

We chose very simple methods for measuring single unit thresholds and combining the individual unit responses into population thresholds. For the single unit thresholds, it is possible that different methods that incorporated spike times (including, but not limited to, synchronized rate, which we discussed in Chapter 2) or utilized more sophisticated methods, such as comparing the responses to some kind of signal template, might change the results presented in this thesis. These ideas are important because, for example, had we used a higher signal level, detection methods based on spike timing may have been essential because the neural response rates might have been saturated. On the other hand, in the case when the excitatory signal response is masked via suppression, the individual thresholds are unlikely to change regardless of the threshold estimation method used; this fact is especially important because these thresholds are usually the most sensitive and would likely remain so regardless of the threshold method. However, when both the signal and masker are excitatory (which, for our methods, gave the worst thresholds), using more sophisticated signal-detection methods may improve the thresholds, as shown by the use of synchronized rate. A visual inspection of the dot rasters generally shows that the thresholds when both the signal and masker are excitatory will be worse than those in the case of an excitatory signal and suppressive masker, although the magnitude of the threshold difference could change. It is less clear that the thresholds for the case

when the noise is excitatory and the signal is suppressive would always be worse than the reverse condition if different methods were used, especially in the cases when the thresholds for these two conditions are within a few dB of each other. It is also not immediately obvious how the population thresholds would change if these more sophisticated methods were used to determine individual unit thresholds. However, population thresholds may be less sensitive to these effects because they are largely determined by the case when the signal was excitatory and the masker was suppressive.

It is also not clear how using a more sophisticated method for determining the population thresholds would change these thresholds. We used the most sensitive unit in our population as our estimate of the population threshold, what is called the “lower envelope principle” by Parker and Newsome (1988). As Parker and Newsome (1988) discuss extensively, the main alternative to the lower envelope principle is to assume that the central processor can pool the individual neuron responses across the population. Such a combined response would result in at least as good as or better population thresholds, assuming a reasonable processing scheme, because the central processor would have access to the best unit as well as all of the others. However, pooling the unit responses was not required in this case to achieve thresholds that were comparable in both magnitude and shape to human behavioral thresholds.

Neural Implementation of Mechanisms

Here we discuss possible neural implementations for the signal processing pathways in the envelope-processor model. In this section, we suggest specific physiological experiments that could be done to investigate the neural implementation. Other future work involving psychophysics and electronic devices is described in a later section.

First, it seems likely that the IC units’ ITD sensitivity is derived from inputs from the MSO that are thought to act as interaural cross-correlators. But at what stage is the independent processing of sound pressure level (SPL) and ITD combined? There are only a few obvious possibilities: 1) the SPL processing also occurs in the MSO, and there are mechanisms in the MSO which make the ITD and SPL sensitivity essentially

independent; or 2) the MSO responses depend on SPL in a way that is dependent on the ITD processing, and subsequent processing removes the dependence, either through some fairly complex processing or by providing a strong parallel input that swamps the MSO's SPL dependence. It does not seem possible to distinguish between these two choices from our data or from previous studies. However, one example in Goldberg and Brown (1969) shows that the responses to ITD and SPL may be independent at the level of the MSO. The most obvious way to shed light on this question is to record from more units in the MSO, which is technically difficult.

How, then, is the modulation sensitivity implemented? The first neurons in the auditory pathway, the auditory nerve fibers, show sensitivity to envelope modulation, presumably due to neural adaptation (Delgutte et al., 1998). However, the fibers' sensitivity to amplitude modulation is shown through changes in synchronization to the signal envelope (as measured by their tMTF) as opposed to changes in firing rate (as measured by their rMTF). Similarly, the primary-like units in the cochlear nucleus, which provide inputs to the MSO, show low-pass or weakly bandpass tMTFs, but flat rMTFs (Rhode and Greenberg, 1994; Kim et al., 1990). The envelope-processor model, as we described it, requires that the rate, rather than the synchrony, change with modulation, and studies have shown that units in the IC have bandpass rMTFs (Langner and Schreiner, 1988; Krishna and Semple, 2000). A model by Hewitt and Meddis (1994) suggests that having the cells in the CN converge on coincidence detectors could transform the tMTFs into rMTFs; these researchers suggest that the coincidence could occur in the IC. It is possible, however, that for the low-frequency units, this transformation occurs in the binaural coincidence detectors in the MSO, rather than in the IC. Because primary-like units in the CN often have low-pass, instead of highpass or bandpass tMTFs as required by the envelope-processor model, further refinements to the MTFs are likely to be required. One possibility, discussed in Chapter 3, is that the direct, excitatory path from the MSO is combined with an inhibitory input from ITD-sensitive units in the dorsal nucleus of the lateral lemniscus (DNLL) to create a temporal differentiator, which would give a bandpass MTF. Our data cannot determine which of

these mechanisms (or some other mechanisms altogether) is the true neural implementation of the modulation filter. On the other hand, our data does almost rule out the possibility that the modulation filter is due to a parallel, monaural input to the IC, such as one coming directly from the cochlear nucleus. This possibility seems improbable because the change in rate with 40-Hz modulation is very sensitive to changes in signal azimuth, and an input from a monaural pathway would not be very sensitive to azimuth at low frequencies. Consequently, the modulation filter appears to occur somewhere along the binaural pathway through the MSO and IC and is likely to be a cascade of filters that refine the MTF throughout this pathway.

Further physiological experiments could elucidate the nature of the modulation filter required by the envelope-processor model. The mechanisms involved in creating the MTFs for the units in the pathway through the MSO could be specifically targeted in future physiological experiments. As mentioned above, the MTFs for the first stages, the ANF and the primary-like units in the cochlear nucleus, have already been studied. Further recordings from low-frequency, ITD-sensitive units in the MSO, DNLL, and the IC could then determine how the rMTFs in the IC emerge through this ascending ITD-sensitive pathway. Additionally, blocking the inhibitory inputs to the IC cells in a manner similar to that used by Caspary et al. (2002) would specifically test the hypothesis that delayed inhibitory inputs to the IC refine the rMTFs.

The Central Processor

How, then, does the “central processor” process the information coming from the IC? To generalize the envelope-processor model, we imagine that the bandpass signals from the auditory nerve pass through the cross-correlators in the MSO, which are essentially acting to enhance information from a particular interaural delay while suppressing information from another. The outputs of the cross-correlators could then be passed to an array of modulation filters. (However, as a cautionary note, it is not clear that there really is modulation filter bank because the MTFs at the level of the IC have a limited range of best modulation frequencies.) Such a processing scheme would seem to be advantageous

in that it separates sounds according to their location and envelope, which are both important features for auditory streaming. The question, then, is how does the central processor recombine this information to detect the signal? Because the most sensitive units seem to be the ones with peaks in their rate-azimuth functions at the signal location and with increased sensitivity to the modulation rate of the signal, the central processor's job may simply be to focus on these most sensitive units: essentially, the central processor could pay attention to the units that were tuned to the signal's location and repetition rate and ignore the others, a remarkably intuitive scheme. This scheme, however, only works if the listener knows what and where the signal is, as well as when the signal could be present, which was true for our psychophysical experiments. Furthermore, our calculations based on the average rate would make the central processor's job easier—the processor would only need to count the output spikes of these most sensitive units. However, as discussed above, it is possible that better thresholds may be obtained through the use of template-matching procedures, for example. Whether the central processor can use these more sophisticated techniques is unclear, especially since the rate-based population responses gave us reasonable predictions of psychophysical behavior.

Psychophysics

It would also be interesting to test whether modulation sensitivity and binaural processing interact improve signal detection in human behavior. It is possible that such interaction could affect psychophysical results. For example, one could compare the detection thresholds of low-frequency pure tones and amplitude-modulated tones for different signal and masker locations. In this case, one would expect that, if the binaural and envelope processors were interacting, the amplitude-modulated (AM), binaural conditions would yield better thresholds than the improvement seen by just the binaural or just the AM conditions. For this experiment, one would need to be cautious about equalizing the signal level in both cases and keeping the sidebands for the AM tone inside the critical band; the carrier for the AM tone should match the frequency of the

pure tone, and both should be low frequency (below 1 kHz). Interestingly, the envelope-processor model predictions based on the observed IC responses suggest that AM would provide the most advantage when the signal and masker are spatially separated.

Additionally, in Chapter 4, we showed that the population thresholds were similar to the psychophysical thresholds when there was an inverse relationship between each unit's ITD and CF. One could explore this idea more thoroughly by using a low-frequency, narrowband signal in our spatial release from masking paradigm. Based on Figure 6 in Chapter 4, different frequency bands would be expected to give different population thresholds if the inverse relationship held. In particular, the thresholds for signal frequencies above 1 kHz would be predicted to give very different thresholds for the signal at $+90^\circ$ depending on whether or not the listener could use units with high CFs and large best ITDs.

Hearing Aids

This work suggests that binaural cross-correlation and sensitivity to modulation are important for a normal-hearing listener's ability to detect sounds in noise. Presumably, this neural processing is also involved in a listener's ability to understand speech in noise. The auditory nerve model used in the envelope-processor model (Zhang et al., 2001) has parameter settings for an impaired periphery so that the model responses for a normal and impaired periphery could be compared. One possible change with hearing impairment is that amplitude modulation encoded in the normal auditory nerve may become distorted with sensorineural hearing loss. It is possible that hearing aids could be designed to compensate for any changes seen with the impaired model. At the very least, an attempt to consider this neural processing could improve the performance of hearing aid users in noise. For example, this study re-emphasizes the idea that the phase responses of binaural hearing aids should be matched to improve a listener's chances to use the binaural system to improve speech intelligibility. Also, hearing aid designers should be cautious with automatic gain control, which could interfere with the envelope of the

incoming sounds and interrupt the listener's ability to use amplitude modulation as a cue for signal detection and speech recognition.

Automatic speech recognition

Preprocessing speech in noise in a way similar to the model described here could improve the performance of automatic speech recognizers in noise. Essentially, the output of two microphones would be cross-correlated and then passed through a bank of modulation filters. The speech recognition system would then process the output of the modulation filters. Although it may not always be possible to cross-correlate the inputs from two speakers (for example, in telephone applications), the modulation filter bank may still give improved performance in noise. One example of improved automatic speech recognition in noise using a modulation filter bank was described by Kleinschmidt et al. (2001); the addition of a modulation filter bank did improve the recognition of German digits in noise.

Final Conclusion

Overall, we suggest that modulation sensitivity and interaural cross-correlation are used cooperatively to improve signal detection. This idea, which seems fairly obvious once stated, appears to be a novel one. Consequently, we are eager to see how ideas from these two usually disparate fields of auditory research may complement each other.

References

- Adams, J. (1995) Cytochemical mapping of the inferior colliculus. *Assoc Res Otolaryngol Abstr* 160, 1995.
- Boudreau, J.C. and Tsuchitani, C. (1970) Cat superior olive S-segment cell discharge to tonal stimulation. *Contrib Sens Physiol* 4:143-213.
- Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (2002) Precise inhibition is essential for microsecond interaural time difference coding. *Nature* 30:543-7.
- Bronkhorst A.W. and Plomp, R. (1988) The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J Acoust Soc Am* 83:1508-1516.
- Brown, T. J. (2000). Characterization of Acoustic Head-Related Transfer Functions for Nearby Sources. M.S. Thesis. Electrical Engineering and Computer Science. Cambridge, MA, Massachusetts Institute of Technology.
- Cai, H., Carney, L.H., Colburn, H.S. (1998) A model for binaural response properties of inferior colliculus neurons. I. A model with interaural time difference-sensitive excitatory and inhibitory inputs. *J Acoust Soc Am*. 103:475-93.
- Casparry, D.M., Palombi, P.S., and Hughes, L.F. (2002) GABAergic inputs shape responses to amplitude modulated stimuli in the inferior colliculus. *Hear Res*. 168:163-73.
- Colburn, H. S. (1977a) Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. *J Acoust Soc Am* 61:525-533.
- Colburn, H. S. (1977b) Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. Supplementary material. AIP document no. PAPS and JASMA-610525-98.
- Colburn, H.S. (1973) Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination. *J Acoust Soc Am* 54:1458-1470.
- Colburn, H.S. (1977a) Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. *J Acoust Soc Am* 61:525-533.
- Colburn, H.S. (1977b) Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. Supplementary material. AIP document no. PAPS and JASMA-610525-98.
- Colburn, H.S., and Durlach, N.I. (1978) Models of binaural interaction. In E.C. Carterette and M.P. Friedman (Eds), *Handbook of Perception*, New York,: Academic Press, pp 467-518.
- Delgutte, B., Hammond. B.M., and Cariani, P.A. (1998) Neural coding of the temporal envelope of speech: Relation to modulation transfer functions. In *Psychophysical and Physiological Advances in Hearing*, A.R. Palmer, A. Rees, A.Q Summerfield, and R. Meddis (eds). London, Whurr: 595-603.
- Durlach N.I. and Colburn H.S. (1978) Binaural Phenomena. In E.C. Carterette and M.P. Friedman (Eds), *Handbook of Perception*, New York,: Academic Press, pp 365-466.

- Gilkey, R.H. and Good, M.D. (1995) Effects of frequency on free-field masking. *Hum Factors* 37:835-843.
- Goldberg, J.M., and Brown, P.B. (1969) Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J Neurophysiol* 32: 613-636.
- Good, M.D., Gilkey, R.H., and Ball, J.M. (1997) The relation between detection in noise and localization in noise in the free field. In R.H. Gilkey and T.R. Anderson (Eds), *Binaural and Spatial Hearing in Real and Virtual Environments*. Lawrence Erlbaum Associates, Mahwah, N.J, pp 349–376.
- Guinan J.J., Jr., Norris B.E., and Guinan, S.S. (1972b) Single Auditory Units in the Superior Olivary Complex II: Locations of Unit Categories and Tonotopic Organization. *Intern J Neuroscience* 4:147-166.
- Hancock, K.E. and Delgutte, B. (2003) Tuning to interaural time difference in the cat inferior colliculus: Dependence on characteristic frequency. *Abstr Assoc Res Otolaryngol* 26, 2003.
- Henkel, C.K. and Spangler, K.M. (1983) Organization of the efferent projections of the medial superior olivary nucleus in the cat as revealed by HRP and autoradiographic tracing methods. *J Comp Neurol* 221: 416–428.
- Hewitt, M.J. and Meddis, R. (1994) A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus. *J Acoust Soc Am*. 95:2145-59.
- Irvine, D. R. F. (1992) Physiology of the auditory brainstem. In A. N. Popper and R. R. Fay (Eds), *The Mammalian Auditory Pathway: Neurophysiology*. Springer-Verlag, New York, pp 169-183.
- Jeffress, L. (1948) A place theory of sound localization. *J. Comp Physiol Psychol* 41:35-39.
- Jeffress, L.A., Blodgett, H.C., Sandel, T.T., and Wood, C.L., I (1956) Masking of Tonal Signals. *J Acoust Soc Am* 28:416-426.
- Jiang, D., McAlpine, D., and Palmer, A.R. (1997a) Detectability index measures of binaural masking level difference across populations of inferior colliculus neurons. *J Neurosci* 17, 9331-9339.
- Jiang, D., McAlpine, D., and Palmer, A.R. (1997b) Responses of neurons in the inferior colliculus to binaural masking level difference stimuli measured by rate-versus-level functions. *J Neurophysiol* 77:3085-3106.
- Johannesma, P.I.M. (1972) The pre-response stimulus ensemble of neurons in the cochlear nucleus. In B.L. Cardozo, E. de Boer, and R. Plomp (Eds), *IPO Symposium on Hearing Theory*. IPO, Eindhoven, The Netherlands, pp. 58-69.
- Joris, P.X., Heijden, M., Louage, D. van de Sande, B. and van Kerckhoven, C. (2003) Dependence of binaural and cochlear “best delays” on characteristic frequency. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models*.
- Kiang, N.Y. and Moxon, E.C. (1974) Tails of tuning curves of auditory-nerve fibers. *J Acoust Soc Am* 55, 620–30.
- Kim, D.O. and Molnar, C.E. (1979) A population study of cochlear nerve fibers:

- comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones. *J Neurophysiol*, 42, 16-30.
- Kim, D.O., Sirianni, J.G., and Chang, S.O. (1990) Responses of DCN-PVCN neurons and auditory nerve fibers in unanesthetized decerebrate cats to AM and pure tones: analysis with autocorrelation/power-spectrum. *Hear Res.* 45:95-113.
- Kleinschmidt, M., Tchorz, J., and Kollmeier, B. (2001) Combining Speech Enhancement and Auditory Feature Extraction for Robust Speech Recognition. *Speech Communication* 24:75-91.
- Krishna, B.S. and Semple, M.N. (2000) Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J Neurophysiol.* 84:255-73.
- Kuwada, S. and Yin, T.C. (1983) Binaural interaction in low-frequency neurons in inferior colliculus of the cat. I. Effects of long interaural delays, intensity, and repetition rate on interaural delay function. *J Neurophysiol* 50:981-999.
- Kuwada, S., Batra, R., and Stanford, T.R. (1989) Monaural and binaural response properties of neurons in the inferior colliculus of the rabbit: effects of sodium pentobarbital. *J Neurophysiol.* 61:269-82.
- Kuwada, S., Batra, R., Yin, T.C., Oliver, D.L., Haberly, L.B., and Stanford, T.R. (1997) Intracellular recordings in response to monaural and binaural stimulation of neurons in the inferior colliculus of the cat. *J Neurosci* 17:7565-81.
- Lane, C.C., Delgutte, B., and Colburn, H.S. (2003a) A population of ITD sensitive units in the cat inferior colliculus shows correlates of spatial release from masking. *Abstr Assoc Res Otolaryngol Abstr* 960, 2003.
- Lane, C.C., Kopco, N., Delgutte, B., Shinn-Cunningham, B.G., and Colburn, H.S. (2003b). A cat's cocktail party: Psychophysical, neurophysiological and computational studies of spatial release from masking. To appear in Pressnitzer, D., de Cheveigne, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models*.
- Langner, G. and Schreiner, C.E. (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J Neurophysiol.* 60:1799-822.
- Levitt, H. and Rabiner, L. R. (1967) Predicting Binaural Gain in Intelligibility and Release from Masking for Speech. *J Acoust Soc Am* 42:820-829.
- Litovsky, R.Y. and Delgutte, B. (2002) Neural correlates of the precedence effect in the inferior colliculus: Effect of localization cues. *J Neurophysiol* 87, 976-994.
- Litovsky, R.Y., Fligor, B.J., and Tramo, M.J. (2002) Functional role of the human inferior colliculus in binaural hearing. *Hear Res* 165:177-88.
- McAlpine D., Jiang D., and Palmer A.R. (2001) A neural code for low-frequency sound localization in mammals. *Nat Neurosci* 4:396-401.
- McAlpine D., Jiang D., and Palmer A.R. (2001) A neural code for low-frequency sound localization in mammals. *Nat Neurosci* 4:396-401.
- McAlpine, D. and Palmer, A.R. (2002) Binaural bandwidths of inferior colliculus neurones measured using interaurally-delayed noise. *Assoc Res Otolaryngol Abstr* 155, 2002.

- McAlpine, D. and Palmer, A.R. (2002) Binaural bandwidths of inferior colliculus neurones measured using interaurally-delayed noise. *Assoc Res Otolaryngol Abstr* 155, 2002.
- Merzenich, M.M. and Reid M.D. (1998) Representation of the cochlea within the inferior colliculus of the cat. *Brain Res* 77, 397–415, 1974.
- Morse, P.M. and Ingard, K.U. (1968) *Theoretical Acoustics*. New York: McGraw-Hill, 1968.
- Müller-Preuss, P., Flachskamm, C., and Bieser, A. (1994) Neural encoding of amplitude modulation within the auditory midbrain of squirrel monkeys. *Hear Res.* 80:197-208.
- Musican, A.D., Chan, J.C., and Hind, J.E. (1990) Direction-dependent spectral properties of cat external ear: new data and cross-species comparisons. *J Acoust Soc Am* 87, 757-781.
- Oliver, D.L. and Morest, D.K. (1984) The central nucleus of the inferior colliculus in the cat. *J Comp Neurol* 222:237-264.
- Oliver, D.L., Beckius, G.E., Bishop, D.C., and Kuwada, S. (1997) Simultaneous anterograde labeling of axonal layers from lateral superior olive and dorsal cochlear nucleus in the inferior colliculus of cat. *J Comp Neurol* 382:215-229.
- Palmer, A.R., Jiang, D., and McAlpine, D. (1999) Desynchronizing responses to correlated noise: A mechanism for binaural masking level differences at the inferior colliculus. *J Neurophysiol* 81:722-734.
- Palmer, A.R., Jiang, D., and McAlpine, D. (2000) Neural responses in the inferior colliculus to binaural masking level differences created by inverting the noise in one ear. *J Neurophysiol* 84:844-52.
- Parker, A.J., and Newsome, W.T. (1998) Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci.* 21:227-77.
- Pena, J.L. and Konishi, M. (2001) Auditory spatial receptive fields created by multiplication. *Science.* 292:249-52.
- Rees, A. and Moller, A.R. (1987) Stimulus properties influencing the responses of inferior colliculus neurons to amplitude-modulated sounds. *Hear Res.* 27: 129-43.
- Rhode, W.S. and Greenberg, S. (1994) Encoding of amplitude modulation in the cochlear nucleus of the cat. *J Neurophysiol.* 71:1797-825.
- Rosowski, J.J., Carney, L.H., and Peake, W.T. (1988) The radiation impedance of the external ear of cat: measurements and applications. *J Acoust Soc Am* 84, 1695-708.
- Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D.R. (1991) Free-field release from masking. *J Acoust Soc Am* 90:1355-1370.
- Sterbing, S.J., D'Angelo, W.R., Ostapoff, E.M., and Kuwada, S. (2003) Effects of amplitude modulation on the coding of interaural time differences of low-frequency sounds in the inferior colliculus. I. response properties. To appear in Jul 30, *J Neurophysiol*.
- Yang, L., Pollak, G.D., and Resler, C. (1992) GABAergic circuits sharpen tuning curves and modify response properties in the mustache bat inferior colliculus. *J Neurophysiol.* 68:1760-74.
- Yin, T.C. (1994) Physiological correlates of the precedence effect and summing

- localization in the inferior colliculus of the cat. *J Neurosci.* 14:5170-86.
- Yin, T.C. and Chan, J.C. (1990) Interaural time sensitivity in medial superior olive of cat. *J Neurophysiol* 64:465-88.
- Young, E.D., Spirou, G.A., Rice, J.J., and Voigt, H.F. (1992) Neural organization and responses to complex stimuli in the dorsal cochlear nucleus. *Philos Trans R Soc Lond B Biol Sci* 336:407-13.
- Zhang, X., Heinz, M.G., Bruce, I.C., and Carney, L.H. (2001) A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. *J Acoust Soc Am.* 109:648-70.
- Zurek, P. M. (1993) Binaural advantages and directional effects in speech intelligibility. In G.A. Studebaker and I. Hochberg (Eds), *Acoustical factors affecting hearing aid performance*, 2nd Edition. Allyn and Bacon, Boston, pp 255-27
- Zweig, G. (1991) Finding the impedance of the organ of Corti. *J Acoust Soc Am.* 89:1229-54.